

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:07:40 ; Search time 13.02 Seconds
(without alignments)
1470.723 Million cell updates/sec

Title: US-09-425-501-2

Perfect score: 2981

Sequence: 1 MVLKRNIDPRHLCHTALPR.....AVEYSDSEDDSEFDEVWLE 559

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2981	100.0	559	1 Y269 HUMAN	Q92558 homo sapien
2	334	11.2	1567	1 DIAL_MOUSE	Q9J104 mus musculu
3	318.5	10.7	1255	1 DIAL_MOUSE	O08808 mus musculu
4	313.5	10.5	1248	1 DIAL_HUMAN	O60610 homo sapien
5	309	10.4	620	1 EXTN_TOBAC	P13983 nicotiana t
6	297	10.0	520	1 WASP_MOUSE	P70315 mus musculu
7	286.5	9.6	426	1 EXLP_TOBAC	O03211 nicotiana t
8	286.5	9.6	1206	1 FM14_MOUSE	Q05859 mus musculu
9	284	9.5	503	1 WAIP_HUMAN	O43516 homo sapien
10	281	9.4	485	1 SP62_MOUSE	O62203 mus musculu
11	280	9.4	502	1 WASP_HUMAN	P42768 homo sapien
12	275.5	9.2	487	1 EBN2_EBV	P12978 epstein-bar
13	275	9.2	1059	1 CAPU_DROME	Q24120 drosophila
14	273.5	9.2	485	1 SSGP_VOLICA	P21997 volvox cart
15	272.5	9.1	1468	1 FMN1_MOUSE	O05860 mus musculu
16	270.5	9.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
17	269.5	9.0	346	1 PRF1_LYCES	O00451 lycopersico
18	269.5	9.0	1213	1 FMN_CHICK	O05858 gallus gall
19	268.5	9.0	817	1 VRP1_YEAST	P37370 saccharomyc
20	267.5	9.0	2715	1 TRX2_HUMAN	Q9unn6 homo sapien
21	264	8.9	464	1 SP62_HUMAN	Q15428 homo sapien
22	263	8.8	267	1 EXTN_MAIZE	P14918 zea mays (m
23	259.5	8.7	306	1 EXTN_DAUCA	P06599 daucus caro
24	257.5	8.6	534	1 APG_ARATH	P40602 arabidopsis
25	256	8.6	1790	1 SEPA_EMENI	P78821 emericeella
26	253.5	8.5	1183	1 DRPL_RAT	P54258 rattus norv
27	251	8.4	633	1 LAI1_YEAST	Q12446 saccharomyc
28	250	8.4	543	1 VP61_NPVAC	Q03209 autographa
29	248.5	8.3	424	1 EXTN_HUMAN	Q15427 homo sapien
30	248	8.3	283	1 EXTN_SORBI	P24152 sorghum bic
31	247	8.3	279	1 Y091_NPVOP	O10341 orgyia pseu
32	246.5	8.3	1185	1 DRPL_HUMAN	P54259 homo sapien
33	246.5	8.3	3149	1 TEGU_EBV	P03186 epstein-bar

34	245.5	8.2	721	1 YK82_MYCTU	Q10690 mycobacteri
35	243.5	8.2	261	1 PRP2_MOUSE	P05142 mus musculu
36	243.5	8.2	474	1 VP61_NPVOP	O10270 orgyia pseu
37	243	8.2	582	1 MNT_HUMAN	O09583 homo sapien
38	242.5	8.1	224	1 Y091_NPVAC	P41479 autographa
39	242	8.1	1794	1 YAV1_SCHPO	Q10172 schizosacch
40	241	8.1	296	1 PRP3_MOUSE	P05143 mus musculu
41	238.5	8.0	3421	1 TEGU_HSVB	P28955 equine herp
42	237.5	8.0	1083	1 T2D3_HUMAN	O00268 homo sapien
43	236	7.9	707	1 SFPO_HUMAN	P23246 homo sapien
44	235	7.9	1048	1 SRA4_RAT	Q03627 rattus norv
45	235	7.9	3164	1 TEGU_HSV1	P10220 herpes simp

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	559 AA.
Y269_HUMAN				
ID	Y269_HUMAN			
AC	Q92558;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269.			
GN	KIAA0269.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
RX	MEDLINE=97191544; PubMed=9039502;			
RA	Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,			
RA	Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. VI.			
RT	The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by			
RT	analysis of cDNA clones from cell line KG-1 and brain."			
RL	DNA Res. 3:321-323(1996)			
CC	-1- SIMILARITY: TO C.ELEGANS R06C1.B.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: D87459; BAA13399.1;			
KW	Hypothetical protein.			
FT	DOMAIN 278 283			
FT	DOMAIN 322 332			
FT	DOMAIN 348 359			
FT	DOMAIN 369 374			
FT	DOMAIN 424 435			
SQ	SEQUENCE 559 AA; 61652 MW; 44B4527BDB77BC6E CRC64;			

Query Match	100.0%;	Score 2981;	DB 1;	Length 559;
Best Local Similarity	100.0%;	Pred. No. 1.5e-133;		
Matches 559;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MPLVKRNIDPRHLCHTALPRGKNELECVTNISLANIIRQLSSLSKYAEIDFGLFNEAH	60	
Db	1	MPLVKRNIDPRHLCHTALPRGKNELECVTNISLANIIRQLSSLSKYAEIDFGLFNEAH	60	
QY	61	SFSFRVNSLQERVDRLSVSVTQLDPKKEELSLQDIWTKAFRSSTIQDQOLFDRKTLPIP	120	
Db	61	SFSFRVNSLQERVDRLSVSVTQLDPKKEELSLQDIWTKAFRSSTIQDQOLFDRKTLPIP	120	
QY	121	LQETDYDCEQPPPLNLTPTVRDDGKGLKFTYNTPSFFFDLWKEKMLQDTEDEKRRKQK	180	

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 15:04:28 ; Search time 3470.44 Seconds
(without alignments)
11699.624 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 ctctctgcacttcgcat.....aatggttcattttaaagtt 2625

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
11 number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_rod:
49: em_sts:
50: em_sy:
51: em_un:
52: em_v1:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_v11:
59: gb_v12:
60: gb_btg1:
61: gb_btg2:
62: gb_btg3:
63: gb_btg4:
64: gb_btg5:
65: gb_btg6:
66: gb_btg7:
67: gb_btg8:
68: gb_btg9:
69: gb_btg10:
70: gb_btg11:
71: gb_btg12:
72: gb_btg13:
73: gb_btg14:
74: gb_btg15:
75: gb_btg16:
76: gb_btg17:
77: gb_btg18:
78: gb_btg19:
79: gb_btg20:
80: gb_btg21:
81: gb_btg22:
82: gb_btg23:
83: gb_btg24:
84: gb_btg25:
85: gb_pr1:
86: gb_pr2:
87: gb_pr3:
88: gb_pr4:
89: gb_pr5:
90: gb_pr6:
91: gb_pr7:
92: gb_pr8:
93: gb_pr9:
94: gb_rod:
95: gb_rod2:
96: gb_in4:
97: gb_pr10:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	2625	100.0	2625	91	D87459	D87459 Human-mRNA
2	1782.6	87.9	2469	94	AF290877	AF290877 Mus muscu
3	1758	67.0	1758	89	AF134303	AF134303/Homo sapi
C 4	861.4	32.8	151580	82	AL590009	AL590009 Homo sapi
C 5	861.4	32.8	192017	82	AL590549	AL590549 Homo sapi
6	837.4	31.9	213475	87	AC019100	AC019100 Homo sapi
7	751.2	28.6	143812	66	AC021847	AC021847 Homo sapi
8	282.6	10.8	1509	85	AB026543	AB026543 Homo sapi

9	278.2	10.6	1497	85	AB026542	Homo sapi
10	264.6	10.1	104597	92	HS393P12	AB022578 Homo sapi
11	262.4	10.0	1491	89	AF134304	Homo sapi
12	205.6	7.8	1368	89	AF134305	Homo sapi
13	205.6	7.8	4450	85	AB020707	Homo sapi
14	173.4	6.6	143812	66	AC021847	Homo sapi
15	173.4	6.6	192017	82	AL590549	Homo sapi
16	171.8	6.5	100000	91	AP000508	Homo sapi
17	171.8	6.5	236822	91	DB4394	Homo sapi
18	157.4	6.0	46797	85	AC004204	Homo sapi
19	139.4	5.3	167390	86	AF007263	Homo sapi
20	136.4	5.2	2845	5	AF247763	Drosophil
21	136.4	5.2	78507	65	AC019704	Drosophil
22	136.4	5.2	176306	4	AC007186	Drosophil
23	136.4	5.2	276576	4	AE003630	Drosophil
24	135.6	5.2	318930	74	AC073495	Mus muscu
25	131.6	5.0	88412	86	AC005918	Homo sapi
26	127.4	4.9	180508	64	AC017042	Homo sapi
27	126.8	4.8	186487	84	CNS059DS	Homo sapi
28	124	4.7	181179	90	AL161725	Homo sapi
29	122.8	4.7	1934	14	PHGCR1	XO4335.Petunia gri
30	121.8	4.6	167985	75	AC078821	Homo sapi
31	120.4	4.6	162269	85	AC000097	Homo sapi
32	120.4	4.6	188239	86	AC007863	Homo sapi
33	120.4	4.6	174840	86	AC006549	Homo sapi
34	120.4	4.6	179269	86	AC006547	Homo sapi
35	119.6	4.6	152528	64	AC016063	Homo sapi
36	118.8	4.5	184929	88	AC058790	Homo sapi
37	117.4	4.5	168354	93	AC025329	Homo sapi
38	117.2	4.5	41185	93	HSBILIC6	Homo sapi
39	117	4.5	103270	13	ATFN20	AL049849 Homo sapi
40	117	4.5	196247	13	ATFRIN20	AL022140 Arabidops
41	117	4.5	200542	83	CNS01RG3	AL161556 Arabidops
42	116	4.4	62537	78	AC090598	Homo sapi
43	114.2	4.4	97348	94	AF091216	Homo sapi
44	114	4.3	75657	93	HSBIL17C8	AL078472 Homo sapi
45	114	4.3	152081	86	AC007381	Homo sapi

ALIGNMENTS

61	CAACATT	CACGTCCTG	CCCTATA	ACCATTAA	TATTTGAT	TCCCGCTAG	ACTAGT	120						
121	ggagaaat	cagcatgt	tataaac	aactgt	tgatag	actgttg	gagtaaa	gttcagtg	180					
121	GGAGAAAT	CAGCATGT	TAACAA	CACTGTT	GATGAT	AGCTGTT	GAGTAA	GTTGCAGTG	180					
181	aagctatg	gctgcaaa	atcgtt	aaaa	tcttc	aagg	tgaa	cgtgcacaa	ggtta	240				
181	AAGCTAT	GGCTGCA	AAATCG	TAAAA	TCTTCA	AGGTGA	CTGCGCAC	AAAGGTTAA	TCTCA	240				
241	agatccgc	tagt	gaaa	aaacat	catc	ctag	gcact	gtgcc	acacag	actgc	300			
241	AGATCCGC	TAGTG	AAAA	AGAAACAT	CGATC	CTAGG	CACTTGTGCC	ACACAG	CACTGCC	TA	300			
301	gaggcatta	agaatga	ac	ctgga	atgtg	ta	acc	aatatt	ctctt	ggaata	atata	360		
301	GAGGCATT	AAGAAT	GAAC	TGGAA	TGTGT	TAC	CAATAT	TTCCTT	GGCAAA	TATATA	ATTTAG	360		
361	aactaag	tagc	taag	taaat	at	ctg	ta	agata	tatt	tg	gagaa	tatt	420	
361	AAC	TAA	GTAG	G	AG	CTTA	ATAT	TCTG	GAAG	ATAT	TTTGG	AGAA	420	
421	atagtttt	ctct	cagag	tcaact	cat	tg	ca	aga	ac	gtg	gg	acc	gtt	480
421	ATAGT	TTTTCT	TCAG	AGTCA	ACTCAT	TGCA	AGA	ACGTGT	G	ACCG	TTTAT	CTG	TTAG	480
481	ttacacag	cttgat	ccaa	agga	aga	atgt	t	ctt	tg	ca	agata	tata	aa	540
481	TTAC	ACGCTTG	ATCC	AAAG	GAAGAA	TGTCT	TTG	CAAG	ATATA	TAA	CAAT	CAG	GAAG	540
541	ctttccg	aa	gtt	ct	taca	att	ca	ag	ccag	ca	gct	ttt	cg	600
541	CTTCCG	AA	GTTCT	CA	CA	ATTCA	AG	CCAG	CA	GCTTT	TCG	AT	CG	600

541	Db	CTTTCCGAAGTCTACANATCAAGACCAAGCAGCTTTTCGATCGAAGACTTTGGCTATATTC	600
601	Qy	cattacaggagacgtacgattgtttgaaacagctccacacctcaataatactcaactcctt	660
601	-Db	CAATTACAGGAGACGTACGATGTTGTGTGAACGCCTCCACCTCTCAATTACTCACTCCTT	660
661	Qy	atagagatgattgaaagaaggtctgaagttttatataccaactcttcgtatttcttggatc	720
661	Db	ATAGAGATGATGTAAGAAGGCTCTGAAGTTTTATACCAATCCTTTGTTATTTCTTTGATC	720
721	Qy	tatggaaagaaaaattgttgcagatcacagagatataagagaagaaagagaagcaga	780
721	Db	TATGGGAAGAAAAATGTTGCAAGATACAGACGATTAAGAGGAAGAAAGAGGAAGCAGA	780
781	Qy	agcagaaaaatctagatcgtctcatgaaccagaaaaagtgccaagagacacctcatgaca	840
781	Db	AGCAGAAAAATCTAGATCTCTCATGAACCAAGAAAAAGTGCCRAAGACACCTCATGACA	840
841	Qy	ggcggcgaagaatgcgaagactggcccaaggtccagagctggctgaagatgatctaact	900
841	Db	GGCGGCGAATGGCAAGAGCTGGCCCAAGTCCAGAGCTGGCTGAAGATGATGCTAATC	900
901	Qy	tcttataagcatattgaagttgcttaatggcccagcctctcattttgaaacagacctc	960
901	Db	TCCTTACATTAAGCATATTGAAGTTGCTAATGGCCCCAGCCTCTCATTTTGAACAAGACGCTC	960
961	Qy	agacatactgggatcatatggatggatcttactactcttctgccttgcctttagtcaga	1020
961	Db	AGACATACCTGGGATCATATGGATGGATCTTACTCACTTCTGCTTGCATTTAGTCAGA	1020
1021	Qy	tgagtgaagcttctgactagagctgaggaaaggatttagtcagaccacatgaacacctc	1080
1021	Db	TGAGTGAGCTTCTGACTAGAGCTGAGAAAGGATATTAGTCAGACCACATGAACACCTC	1080
1081	Qy	caactccaccaatgcattgagcagagatgaaaaaccatgataccacotgtataagttctg	1140
1081	Db	CACCTCCACCAATGCATGGAGCAGAGATGCAAAACCAGATACCCACCTGTATCAGTTCTG	1140
1141	Qy	ctacaggtttgatagaaaaatcgccctcagtcaccagctcacaggcagaaacacctgtttg	1200
1141	Db	CTACAGGTTTGATAGAAAAATCGCCCTCAGTCACCAAGCTACAGGCAGAACACCTGTGTTG	1200
1201	Qy	tgagcccaactcccacactctccacacactcttccacttgccttgccttgcactctccat	1260
1201	Db	TGAGCCCCACTCCCCACCTCTCTCCACCACTCTTCCATCTGCCTTGTCAACTTCCCTCAT	1260
1261	Qy	taagagcttcaatgacttcaactctctccctccagttacattccccacctccacctccag	1320
1261	Db	TAGAGCTTCAATGACTTCAACTCTCTCCCTCTCCAGTACCTTCCCCACCTCCACCTCCAG	1320
1321	Qy	ccactgctttgcaagctccagcagttaccacacacotccagctctcttcagattgccccctg	1380
1321	Db	CCACTGCTTTGCAAGCTCCAGCAGTACCACCACTCCAGCTCCTCTTTCAGATGCCCCCTG	1380
1381	Qy	gagtttccacccagctctctccaatgacatctctctagtacagccctctccaccag	1440
1381	Db	GAGTTCTTCCACCAGCTCTCTCTCCAATTGCACTCTCTTAGTACAGCCCTCTCCACCAG	1440
1441	Qy	tagctagagctgccagttatgtagactgtaccagttcatccactcccacagggtgaag	1500
1441	Db	TAGCTTAGAGCTGCCCGAGTATGTAGACTGTACCAGTTTCATCCACTCCCAAGGTGAAG	1500
1501	Qy	ttcaggggctgctccaccccaacacgcctctctctgctccacctggattctgacct	1560
1501	Db	TTCAGGGGCTGCTCCACCCCAACACCGCTCTCTTCTGCTTCCACCTGGGATTCGACCAT	1560
1561	Qy	catcaactgtcacagttacagctcttgcatactctccctgggctacatcccaactccat	1620
1561	Db	CATCACTGTCAAGTTACAGCTCTTGTCTCATCTCTCTCTGCGCTACATCCCAACTCCAT	1620
1621	Qy	ctactgccccagggtcccattgttccattaatgctccatctctcccatccacaagttatcc	1680
1621	Db	CTACTGCCCAAGGTCCCATTGTTCCTTCCATTAAATGCTCCCATCTCTCCATCCACAAGTTATCC	1680

Db	1621	CTACTGCCCCAGGTCGCCCATGTTCCTATTAAATGCCCTCCATCTCCTCCATCACAAAGTTATAC	1680
Qy	1681	ctgcttctagccaaagccatccatcaaacctaccctgtaataatcaagtgtatgcaggaggtg	1740
Db	1681	CTGCTTCTGAGCCAAAGCGCATCCATCAACCTACCTGTATATCAGTATGCCAGGAGTG	1740
Qy	1741	tgctactgggaagcaatacacaaaaaggtattccaagtacgcaaaagtagaagagcagcgtgaac	1800
Db	1741	TGCTACTGGAAGCAATACGAAAAAGGTATTTCAGCTACGCAAGTAGAAGAGCAGCGTGAC	1800
Qy	1801	aggaagctaagcaigaacgattgaataaaacgattgttgccaccatccctgtctgcgcgtattg	1860
Db	1801	AGGAAGCTAAGCATGAACCATGTGAACACGATGTGTGCCACCATCTGTCTGCCGTATTG	1860
Qy	1861	ctgttgaataatagtattcggaagatgattcagaatttgatgaagtaagtgttggtgaagt	1920
Db	1861	CTGTTGNATATAGTGAATTCGGAAGATGATTACGAATTTGATCAAGTAGATTGGTGTGGAGT	1920
Qy	1921	aagaaaaatgcattgtataataattacaaaactgaatgcaaatgtctcttctgttggtgcttgt	1980
Db	1921	AAGAAAAATGCATTGTATAAATATTACAAACGTAAATGCCTTTGTGTGGTCTTGT	1980
Qy	1981	tccttgaataatgttgggtcattctagtggttgggtcttcttcttccctataataaatgaccc	2040
Db	1981	TCCCTTGAAAAATGTTGGTGCATTCTAGTGTGTTTGCCTTTCTTTCTTATAATAAATGACCC	2040
Qy	2041	tttctccataacttttgaatttcgaagaaataattagcatacaatttcaaaactaaatgt	2100
Db	2041	TTTTCTCCATAACTTTTGATTTCTAAGGAAATAATTAGCATACATTTCAAACTTAATGT	2100
Qy	2101	ttcacagtggtctatcttttttcccccgtaaaagacctaatttgggtccaataaaaccact	2160
Db	2101	TTTACAGTGGCTATCTTTTTTTTCCCCCTGAAAGAGACTAATTTGGTCATAATAAACCACT	2160
Qy	2161	aagattaagcatgagcagctgtgttagtagtagcagatttcagtttttttgatatctcta	2220
Db	2161	AAGTATTATAGCATGGACAGCTGTGTTAGAGTAGCAGATTCAGTTTTTTTGATATATCTTA	2220
Qy	2221	attgtacttcttgaattttaaatttaagaaagcaactgaaatgaaactcttgaggcca	2280
Db	2221	ATTGTGCTACTTTGTGAATTTTAAATTTAAGAAAGCACTGAATTGAAATCTTGAGGGCA	2280
Qy	2281	gctgtatctactaataagccttattccatttctctgatgttttaaagaagaagaaacactgccc	2340
Db	2281	GCTGTATCTACTAATGAGCCTTATTCCATTTCTCATCTTTTAAAGAGAAACACACTGCC	2340
Qy	2341	ttgattatagaataacactcagaagtagacatttagctttagttagttagttagttagttagt	2400
Db	2341	TTGATTATACGAATACACTCAGAAAAGTACATTTAGCTTTGTAGTGTGAATTCCTCTTAAG	2400
Qy	2401	gaatgcttgaatttttcaatttatttattgttttttttatataacttgccttatttgaatg	2460
Db	2401	GAAAGCTTGAATTTTTTCATTTTATTTTATTTGTTTATATATCTGCTTATTTGAATG	2460
Qy	2461	tttagcagtatcccttcccacttatattgtgtgtgatgtattgttcttgcctctatagga	2520
Db	2461	TTTAGCAGTATCCCTCTCCACTTATATATTTGTGTGATGATGATTTTCTTGCCTATAGGA	2520
Qy	2521	gttaaaaaacttttccatgtaataactctgacttaaacatacatatgaacttacataactg	2580
Db	2521	GTTAAAAACTTTTTCCATGTGAAATACTCTGACTTAAACATACATGTAACTTACATAACTG	2580
Qy	2581	ttaagaataaacagctctgatttaataaatcggttcatttttaaaagtt	2625
Db	2581	TTAAGAATAACAGCTGCTGATTTTAATAAATGGTTTCATTTTAAAGATT	2625
RESULT 2			
AF290877			
LOCUS	AF290877	2469 bp	mRNA
DEFINITION	Mus musculus WAVE-1 mRNA, complete cds.		
ACCESSION	AF290877		
VERSION	AF290877.1 GI:9831545		
		ROD	10-OCT-2000

RESULT	2
AF290877	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 20, 2001, 12:07:40 ; Search time 13.02 Seconds
(without alignments)
1470.723 Million cell updates/sec

File: US-09-425-501-2

Perfect score: 2981

Sequence: 1 MPLVKNIDPRHLCHTALPR.....AVEYSDSEDDSEFDEVDWLE 559

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

al number of hits satisfying chosen parameters: 93435

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2981	100.0	559	1 Y269_HUMAN	Q92558 homo sapien
2	334	11.2	1567	1 FMN2_MOUSE	Q94104 mus musculus
3	318.5	10.7	1255	1 DIAL_MOUSE	O08808 mus musculus
4	313.5	10.5	1248	1 DIAL_HUMAN	O60810 homo sapien
5	309	10.4	620	1 EXTN_TOBAC	P13983 nicotiana t
6	297	10.0	520	1 WAP2_MOUSE	P70315 mus musculus
7	286.5	9.6	426	1 EXLP_TOBAC	Q03211 nicotiana t
8	286.5	9.6	1206	1 FM14_MOUSE	Q05859 mus musculus
9	284	9.5	503	1 WAP1_HUMAN	Q43516 homo sapien
10	281	9.4	485	1 SP62_MOUSE	Q62203 mus musculus
11	280	9.4	502	1 WAP2_HUMAN	P42768 homo sapien
12	275.5	9.2	487	1 EBN2_EBV	P12978 Epstein-barr
13	275	9.2	1059	1 CAPL_DROME	Q24120 drosophila
14	273.5	9.2	485	1 SSGP_VOLCA	P21897 volvox cart
15	272.5	9.1	1468	1 FMN1_MOUSE	Q05860 mus musculus
16	270.5	9.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
17	269.5	9.0	346	1 PRF1_LYCES	Q00451 lycopersico
18	269.5	9.0	1213	1 FMN_CHICK	Q05858 gallus gall
19	268.5	9.0	817	1 VRP1_YEAST	P37370 saccharomyc
20	267.5	9.0	2715	1 TRX2_HUMAN	Q9unn6 homo sapien
21	264	8.9	464	1 SP62_HUMAN	Q15428 homo sapien
22	263	8.8	267	1 EXTN_MAIZE	P14918 zea mays (m
23	259.5	8.7	306	1 EXTN_DAUCA	P06599 daucus caro
24	257.5	8.6	534	1 APG_ARATH	P40602 arabidopsis
25	256	8.6	1790	1 SEPA_EMENI	P78521 americeila
26	253.5	8.5	1183	1 DRPL_RAT	P54258 rattus norv
27	251	8.4	533	1 LA17_YEAST	Q12446 saccharomyc
28	250	8.4	1343	1 VP61_NPVAC	Q03209 autographa
29	248.5	8.3	424	1 SP49_HUMAN	Q15427 homo sapien
30	248	8.3	283	1 EXTN_SORBI	P24152 sorghum bic
31	247	8.3	279	1 Y091_NPVOP	O10341 orgyia pseu
32	246.5	8.3	1185	1 DRPL_HUMAN	P54259 homo sapien
33	246.5	8.3	3149	1 TEGU_EBV	P03186 Epstein-bar

34	245.5	8.2	721	1 YK82_MYCTU	Q10690 mycobacteri
35	243.5	8.2	261	1 PRP2_MOUSE	P05142 mus musculus
36	243.5	8.2	474	1 VP61_NPVOP	O10270 orgyia pseu
37	243	8.2	582	1 MNT_HUMAN	Q99583 homo sapien
38	242.5	8.1	224	1 Y091_NPVAC	P41479 autographa
39	242	8.1	1794	1 YAV1_SCHPO	Q10172 schizosacch
40	241	8.1	296	1 PRP3_MOUSE	P05143 mus musculus
41	238.5	8.0	3421	1 TEGU_HSVB	P28955 equine herp
42	237.5	8.0	1083	1 T2D3_HUMAN	Q00268 homo sapien
43	236	7.9	707	1 SFPQ_HUMAN	P23246 homo sapien
44	235	7.9	1048	1 SRA4_RAT	Q63627 rattus norv
45	235	7.9	3164	1 TEGU_HSV11	P10220 herpes simp

ALIGNMENTS

RESULT 1

ID Y269_HUMAN STANDARD; PRT; 559 AA.

AC Q92558;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269.
GN KIAA0269.

OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;

RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes. VI.
The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
analysis of cDNA clones from cell line KG-1 and brain.";

RL DNA Res. 3:321-329(1996).

CC -1- SIMILARITY: TO C-ELGANS R06C1.B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL; D87459; BAA13399.1;

DR Hypothetical protein.

FT DOMAIN 278 283 POLY-PRO.

FT DOMAIN 322 332 POLY-PRO.

FT DOMAIN 348 359 POLY-PRO.

FT DOMAIN 369 374 POLY-PRO.

FT DOMAIN 424 435 POLY-PRO.

SQ SEQUENCE 559 AA; 61652 MW; 4484527BD57BC6E CRC64;

Query Match 100.0%; Score 2981; DB 1; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.5e-133;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLVKNIDPRHLCHTALPRGKNELECVTNLSLNIQLSLSKYADIFGELFNEAH 60

DB 1 MPLVKNIDPRHLCHTALPRGKNELECVTNLSLNIQLSLSKYADIFGELFNEAH 60

QY 61 SFSFRVNSLQSDVDRLSVSVTOLDPKKEELSQTDMRKAFRSSTIQDQLFDRKTLTP 120

DB 61 SFSFRVNSLQSDVDRLSVSVTOLDPKKEELSQTDMRKAFRSSTIQDQLFDRKTLTP 120

QY 121 LQETDYVCBCPPPLNLTTPYRDDGKGLNFYTNPSYFFDLWKEKMLQDTEDEKREKQK 180

Db 121 LQSTYDVCQPPPLNLTPTDGRGKGLFTNPSYFDLWKEKMLQDTEKREKQK 180
QY 181 QKNLDRPHEPEKVPAPDRRREWKLAQGPPLAEDDANLLKHKEVANGPASHFETRPQ 240
Db 181 QKNLDRPHEPEKVPAPDRRREWKLAQGPPLAEDDANLLKHKEVANGPASHFETRPQ 240
QY 241 TYVDHMGYSLSALPFFSQMSSELLTRAERVLVRHEPHEPPPMHGCAGDAKPIPTCISSA 300
Db 241 TYVDHMGYSLSALPFFSQMSSELLTRAERVLVRHEPHEPPPMHGCAGDAKPIPTCISSA 300
QY 301 TGLIENRQSPATGTPVFTVSTPTPPPPPLPSALSTSLRASMTSTPPPPPPPPPPA 360
Db 301 TGLIENRQSPATGTPVFTVSTPTPPPPPLPSALSTSLRASMTSTPPPPPPPPPPA 360
QY 361 TALQAPVPPPPAPLQIAGVLHPAPPIAPPLVQSPVAPVAPVAPVAPVAPVAPVAPV 420
Db 361 TALQAPVPPPPAPLQIAGVLHPAPPIAPPLVQSPVAPVAPVAPVAPVAPVAPVAPV 420
QY 421 QGLPPPPPPPLPPPIRPSPTVTTALAHPPSGLHPTPTSTAPGPHVPLMPSPSQVIP 480
Db 421 QGLPPPPPPPLPPPIRPSPTVTTALAHPPSGLHPTPTSTAPGPHVPLMPSPSQVIP 480
QY 481 ASEPKRHPSTLPVSDARSVLEAIRKGIQIRKVEEQREQAKHERIENDVATILSRRIA 540
Db 481 ASEPKRHPSTLPVSDARSVLEAIRKGIQIRKVEEQREQAKHERIENDVATILSRRIA 540
QY 541 VEYSDSEDDSEFDEVDWLE 559
Db 541 VEYSDSEDDSEFDEVDWLE 559

RESULT 2
FMN2_MOUSE STANDARD; PRT: 1567 AA.
AC Q9JL04; 2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FORMIN 2.
GN FMN2.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Brain;
CC MEDLINE=20245324; PubMed=10781961;
CC Rader B., Leder P.;
CC Formin-2, a novel formin homology protein of the cappuccino
CC subfamily, is highly expressed in the developing and adult central
CC nervous system.;
CC Mech. Dev. 93:221-231(2000).
CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN THE DEVELOPING
CC AND MATURE CENTRAL NERVOUS SYSTEM.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT EMBRYONIC DAY 9. 5 IN
CC THE DEVELOPING SPINAL CORD AND BRAIN STRUCTURES AND CONTINUES IN
CC NEONATAL AND ADULT BRAIN STRUCTURES INCLUDING THE OLFACTORY BULB,
CC CORTEX, THALAMUS, HYPOTHALAMUS, HIPPOCAMPUS AND CEREBELLUM.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. CAPPUCCINO
CC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

EMBL: AF218940; AAF72883.1;
MGD; MGI:1859252; Fmn2.
InterPro; IPR001265;
InterPro; IPR003104;
Pfam; PF02181; FH2; 1.
PRINTS; PR00828; FORMIN.
KW Developmental protein; Repeat; Coiled coil.
FT DOMAIN 643 683 COILED COIL (POTENTIAL).
FT DOMAIN 735 1113 FH1 (PRO-RICH).
FT DOMAIN 919 1039 11 X 11 AA TANDEM REPEATS OF [MV]-G-I-P-
P-P-P-L-P-G.
REPEAT 919 929 1.
REPEAT 930 940 2.
REPEAT 941 951 3.
REPEAT 952 962 4.
REPEAT 963 973 5.
REPEAT 974 984 6.
REPEAT 985 995 7.
REPEAT 996 1006 8.
REPEAT 1007 1017 9.
REPEAT 1018 1028 10.
REPEAT 1029 1039 11.
FT DOMAIN 1128 1532 FH2.
FT DOMAIN 1408 1444 COILED COIL (POTENTIAL).
FT DOMAIN 48 55 POLY-GLY.
FT DOMAIN 202 207 POLY-GLN.
FT DOMAIN 797 801 POLY-PRO.
FT DOMAIN 861 864 POLY-PRO.
FT DOMAIN 908 917 POLY-PRO.
FT DOMAIN 922 928 POLY-PRO.
FT DOMAIN 933 939 POLY-PRO.
FT DOMAIN 944 950 POLY-PRO.
FT DOMAIN 955 961 POLY-PRO.
FT DOMAIN 966 972 POLY-PRO.
FT DOMAIN 977 983 POLY-PRO.
FT DOMAIN 988 994 POLY-PRO.
FT DOMAIN 999 1005 POLY-PRO.
FT DOMAIN 1010 1016 POLY-PRO.
FT DOMAIN 1021 1027 POLY-PRO.
FT DOMAIN 1032 1038 POLY-PRO.
FT DOMAIN 1043 1046 POLY-PRO.
FT DOMAIN 1054 1060 POLY-PRO.
FT DOMAIN 1065 1072 POLY-PRO.
FT DOMAIN 1077 1080 POLY-PRO.
SQ SEQUENCE 1567 AA; 166268 MW; 8F273B1C88505944 CRC64;

Query Match 11.2%; Score 334; DB 1; Length 1567;
Best Local Similarity 26.7%; Pred. No. 5e-09;
Matches 146; Conservative 53; Mismatches 184; Indels 164; Gaps 28;
QY 87 EEELSQQDITMRKAFRSSTIQ-----DQOLFDRK-----TLPIQLQETV-----125
Db 610 EEFPSPKVDVDEP-KSSIIESPCKCSNGVQVEFVKSEGOATVIOOLETIEDLTKI 667
QY 126 -DVCEOPPPNLTPTDGRGKGLFTNPSYFDLWKEKMLQDTEKREKQKQKN 183
Db 668 AELEKQYPALDIEGPRGLSGLNGLTASADVSLDALVHLGKVAQ-----PPTLEAKS 720
QY 184 LD-RPEPEKV-----PRPHDRRRRWKLAQGPPLAEDDANLLKHKEVANGPASHFET 237
Db 721 IQSPTEGRILTLPPPKAPP-----EGLPGSPAASGESALLTS-----PSGPQTKFCS 770
QY 238 RPQTYVDHMGYSLSALPFFSQMSSEL-LTRAERVLVRHEP-----278
Db 771 EISLIVSPRISVQLDQAQIQSASQLPPPLLGSDQ--GPSQSLHTESETSHESV 828
QY 279 -----PPPPP-----MHCAGDAKPIPTCISATGLIENRQSPA-----T 313
Db 829 SSSFGNCCNVPPAPPPLPCTESSSEFMPLGMAIPPPCLSDITVPALPSTALQFSLNQ 888
QY 314 GRTPVFTVSTPP-----PPPPPLPSALSTSLRASMTSTPPPP-----VPPPPP 357

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

QM nucleic: nucleic search, using sw model

Run on: June 20, 2001, 15:04:28 ; Search time 3470.44 Seconds
(without alignments)
11699.624 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 cttctctgcacttcggat.....aatgttcattttaaaagt 2625

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database GenEmbl:

1:	gb_ba1.*
2:	gb_ba2.*
3:	gb_ba3.*
4:	gb_in1.*
5:	gb_in2.*
6:	gb_in3.*
7:	gb_cm.*
8:	gb_ov.*
9:	gb_pat1.*
10:	gb_pat2.*
11:	gb_ph.*
12:	gb_pl1.*
13:	gb_pl2.*
14:	gb_pl3.*
15:	gb_pl4.*
16:	em_ba1.*
17:	em_ba2.*
18:	em_ba3.*
19:	em_htgo_hum.*
20:	em_htgo_inv.*
21:	em_htgo_rod.*
22:	em_htg_hum1.*
23:	em_htg_hum2.*
24:	em_htg_hum3.*
25:	em_htg_hum4.*
26:	em_htg_hum5.*
27:	em_htg_hum6.*
28:	em_htg_hum7.*
29:	em_htg_hum8.*
30:	em_htg_inv1.*
31:	em_htg_inv2.*
32:	em_htg_other.*
33:	em_htg_rod.*
34:	em_hum1.*
35:	em_hum2.*
36:	em_hum3.*
37:	em_hum4.*
38:	em_hum5.*
39:	em_hum6.*
40:	em_hum7.*
41:	em_hum8.*
42:	em_cm.*
43:	em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vl.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vl.*
59: gb_vl2.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_r01.*
95: gb_r02.*
96: gb_r04.*
97: gb_r10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2625	100.0	2625	91	D87459 Human mRNA
2	1782.6	67.9	2469	94	AF290877 Mus muscu
3	1758	67.0	1758	89	AF134303 Homo sapi
4	861.4	32.8	51580	82	AL590009 Homo sapi
5	861.4	32.8	192017	82	AL590549 Homo sapi
6	837.4	31.9	213475	87	AC019100 Homo sapi
7	751.2	28.6	143812	66	AC021847 Homo sapi
8	282.6	10.8	1509	85	AB026543 Homo sapi

DR	EMBL; AF218940; AAF72883.1;
DR	MGD; MGI:1859252; Fmn2.
DR	InterPro; IPR001265;
DR	InterPro; IPR003104;
DR	Pfam; PF02181; FH2; 1.
DR	PRINTS; PR00828; FORMIN.
DR	Developmental protein; Repeat; Coiled coil.
KW	DOMAIN 643 693
FT	DOMAIN 735 1113
FT	DOMAIN 919 1039
FT	
FT	

FT	DOMAIN	944	950	POLY-PRO.
FT	DOMAIN	955	961	POLY-PRO.
FT	DOMAIN	966	972	POLY-PRO.
FT	DOMAIN	977	983	POLY-PRO.
FT	DOMAIN	988	994	POLY-PRO.
FT	DOMAIN	999	1005	POLY-PRO.
FT	DOMAIN	1010	1016	POLY-PRO.
FT	DOMAIN	1021	1027	POLY-PRO.
FT	DOMAIN	1032	1038	POLY-PRO.
FT	DOMAIN	1043	1046	POLY-PRO.
FT	DOMAIN	1054	1060	POLY-PRO.
FT	DOMAIN	1065	1072	POLY-PRO.
FT	DOMAIN	1077	1080	POLY-PRO.
SO	SEQUENCE	1567	AA: 166268	MM: 8F273B1C88505944 CRC64:

	Query Watch	11.2%; Score 334; DB 1; Length 1567;
	Best Local Similarity	26.7%; Pred. No. 5e-09;
	Matches 146; Conservative 53; Mismatches 184; Indels 164; Gaps	
QY	87 EEELSLODITMRKAFRSSIQ-----DQOLFPRK-----TLPIQLQET-----	125
	: : : : : :	
Db	610 EEPSPKDVDTPE--KSSILESPKKCSNGVQVFVKSEGOATVQQLEQTIEDLRTKI	667
QY	126 -DYCEOPPPLNLTTRYDDGKE-GLAFYNPVSFFDLWKEKMLQDTEKKRKKOKXN	183
	: : : : : :	
Db	668 AELEKOTPALDEGGRLGLENGLTASADVSDALVLHGKVQA-----PPRTLKAS	720
QY	184 LD-RPHEPEKV----PRAPHDRRRWEQKLQAQPGLAEADANLLHKHIEVANGPSHFET	237
	: : : : : :	
Db	721 IQSPTEGRILLTPPKAPP-----EGLPGSAAAGESALLTS---PSGPQKFCS	770
QY	238 RPTQYVDHMDGSYSLSALPFSQMSEL-----LTPRAERVLVLRPHEP-----	278
	: : : : : :	
Db	771 EISLIYSPRAISVOLDAQTQSOLPPPPPILGDSQ--GOPSQPSLHTESTSHSVS	828
QY	279 -----PPPPP-----MHGAGDAKPICPICISSATGLIENRPQSA-----	313
	: : : : : :	
Db	829 ISSFGNNCNVPAPPLPCTESSFMFGLGNAPPPPCLLSDITVPALPSTAPALQFSNLQ	886
QY	314 GRTPVGVSPPPP-----PPPPPLPSALSTSISRASMTSTPPPP-----VPPPPP	357

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:07:40 ; Search time 13.02 Seconds
(without alignments)
1470.723 Million cell updates/sec

Title: US-09-425-501-2

Perfect score: 2981

Sequence: 1 MPLVKRNIDPRHLCHTALPR.....AVEYSDSEDDSEFDEVDMLE 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2681	100.0	559	1 Y269_HUMAN	Q92558 homo sapien
2	334	11.2	1567	1 FMN2_MOUSE	Q9J104 mus musculu
3	318.5	10.7	1255	1 DIAL_MOUSE	O08808 mus musculu
4	313.5	10.5	1248	1 DIAL_HUMAN	O06010 homo sapien
5	309	10.4	620	1 EXTN_TOBAC	P13983 nicotiana t
6	297	10.0	520	1 WASP_MOUSE	P70315 mus musculu
7	286.5	9.6	426	1 EXLP_TOBAC	Q03211 nicotiana t
8	286.5	9.6	1206	1 FM14_MOUSE	Q05859 mus musculu
9	284	9.5	503	1 WAP1_HUMAN	O43516 homo sapien
10	281	9.4	485	1 SP62_MOUSE	O62203 mus musculu
11	280	9.4	502	1 WASP_HUMAN	P42768 homo sapien
12	275.5	9.2	487	1 EBN2_EBV	P12978 epstein-bar
13	275	9.2	1059	1 CAPL_DROME	Q24120 drosophila
14	273.5	9.2	485	1 SSGP_VOLCA	P21997 volvox cart
15	272.5	9.1	1468	1 FMN1_MOUSE	Q05860 mus musculu
16	270.5	9.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
17	269.5	9.0	346	1 PRF1_LYCES	Q00451 lycopersico
18	269.5	9.0	1213	1 FMN_CHICK	Q05858 gallus gall
19	268.5	9.0	817	1 VRP1_YEAST	P37370 saccharomyc
20	267.5	9.0	2715	1 TRX2_HUMAN	Q9unn6 homo sapien
21	264	8.9	464	1 SP62_HUMAN	Q15428 homo sapien
22	263	8.8	267	1 EXTN_MAIZE	P14918 zea mays (m
23	259.5	8.7	306	1 EXTN_DAUCA	P06599 daucus caro
24	257.5	8.6	534	1 APG_ARATH	P40602 arabidopsis
25	256	8.6	1790	1 SEPA_EMENI	P78621 emericeila
26	253.5	8.5	1183	1 DRPL_RAT	P54258 rattus norv
27	251	8.4	633	1 LAI1_YEAST	Q12446 saccharomyc
28	250	8.4	543	1 VP61_NPVAC	Q03209 autographa
29	248.5	8.3	424	1 SP49_HUMAN	Q15427 homo sapien
30	248	8.3	283	1 EXTN_SORBI	P24152 sorghum bic
31	247	8.3	279	1 Y091_NPVOP	O10341 orgyia pseu
32	246.5	8.3	1185	1 DRPL_HUMAN	P54259 homo sapien
33	246.5	8.3	3149	1 TEGU_EBV	P03186 epstein-bar

RESULT 1
Y269_HUMAN
ID Y269_HUMAN STANDARD; PRT; 559 AA.
AC Q92558;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269.
GN KIAA0269.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT *Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.*;
RL DNA Res. 3:321-329(1996)
CC 1- SIMILARITY: TO C-TERMINUS R06C1.B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; D87459; BAAL3399.1;
KW Hypothetical protein.
FT DOMAIN 278 283 POLY-PRO.
FT DOMAIN 322 332 POLY-PRO.
FT DOMAIN 348 359 POLY-PRO.
FT DOMAIN 369 374 POLY-PRO.
FT DOMAIN 424 435 POLY-PRO.
SQ SEQUENCE 559 AA; 61652 MW; 44B4527BDB77BC6E CRC64;

Query Match 100.0%; Score 2981; DB 1; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.5e-133;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLVKRNIDPRHLCHTALPRGKNEECVTNISIANIROLSSLSKYAEDIFGELFNEAH 60
Db 1 MPLVKRNIDPRHLCHTALPRGKNEECVTNISIANIROLSSLSKYAEDIFGELFNEAH 60
QY 61 SFSFRNLSQERYDRLSVSVTQLDPKKEELSLODITMRKAFRSTTQDQQLDFDKTLPIP 120
Db 61 SFSFRNLSQERYDRLSVSVTQLDPKKEELSLODITMRKAFRSTTQDQQLDFDKTLPIP 120
QY 121 LQETDYDCEOPPLNLITPYRDDGKGLKFTYTPNSYFFDLWKEMLODTEDEKRRKQK 180

Db 889 GEMLPAPPPPPPLGVLGVPVPPPPAPPL-----PGMGIPPPPPPLPGMGIPPPPPPLPGM 941
QY 358 --PPATALQAPVPPPPAPPLQIAPGLVHPAPPIAPPLVQPPVVARAAAPVCETVPHPL 415
Db 942 GIPPPPPPLGVPVPPPPPL-----PGVGIPPPPPPLGVPVPPPPPL-----PGVGIPPPPPPL 993
QY 416 PGEVQGLPPP-----PPPPPPPLPGGIRSPSPVTVTALAHPP-----SGLHPTPSTAP 463
Db 994 PG---VGIPPPPPPLGVPVPPPPPLGVPVPPPPPLPGMGIPPPPPPLPGSGI-PPPPALP 1049
QY 464 GPHVPLMP-----SPSQVTPASEPKRH-----PSTLPVSDARSVLL 502
Db 1050 GVAIPPPPPPLGVPVPPPPPPGAGIPPPPPPLPGSGPPSHSQVGSSTLPAAPQCGGLF 1109
QY 503 EAIRKGI 509
Db 1110 PPLTGL 1116

RESULT 3
DIAL_MOUSE
ID DIAL_MOUSE STANDARD; PRT; 1255 AA.
AC O08808;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 1 (DIAPHANOUS-RELATED FORMIN 1) (DRFL)
DE (MDIAL) (P140MDIA).
GN DIAPH1 OR DIAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357293; PubMed=9214622;
RA Watanabe N., Madaule P., Reid T., Ishizaki T., Watanabe G.,
RA Kazizuka A., Saito Y., Nakao K., Jockusch B.M., Narumiya S.;
RT "p140mdia, a mammalian homolog of Drosophila diaphanous, is a target
RT protein for Rho small GTPase and is a ligand for profilin.";
RN EMBO J. 16:3044-3056(1997).
RN [2]
RP FUNCTION.
RX MEDLINE=20142655; PubMed=10678165;
RA Tomimaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,
RA Alberts A.S.;
RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase
RT signaling";
RL Mol. Cell 5:13-25(2000).
CC -|- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS.
CC -|- SUBCELLULAR LOCATION: MEMBRANE RUFFLES, ESPECIALLY AT THE TIP OF
CC RUFFLES, OF MOTILE CELLS.
CC -|- TISSUE SPECIFICITY: UBIQUITOUS.
CC -|- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION.
CC -|- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -|- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U96963; AAC53280.1;
DR MGD: MGI:1194490; Diapl.
DR InterPro: IPR003104;
DR Pfam: PF02181; FH2; 1.
KW Coiled coil; Repeat.
FT DOMAIN 460 562
FT DOMAIN 63 260
FT DOMAIN 157 457
FT DOMAIN 586 747
FT DOMAIN 1037 1197
FT DOMAIN 1072 1179
FT DOMAIN 1180 1194
FT DOMAIN 1196 1199
FT SEQUENCE 1255 AA; 139343 MW; 09404164873CA7C1 CRC64;
SQ

Query Match 10.7%; Score 318.5; DB 1; Length 1255;
Best Local Similarity 24.2%; Pred. No. 2.2e-08;
Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;
QY 51 IFGELENEAHSPFRVNSLQERVDRLSVTQIDPKKEELSQDITMRKAFRSTTQDOQ 110
Db 308 LINALITPAEELDFRVHI---RSELMRLGLHQVLQELREIENEDMKVLCVFEQ-GDED 363
QY 111 LFDRK-----TLPIPLQETVDYCEQPPPLNILTP---YRDDGREGLKFY- 151
Db 364 FFDLGRLLDDIRMEMDDFGEVFOILLNTVKDSKAEPFLSILQHLNLLNRDYEAFQYK 423
QY 152 -----TNPSY---FFDLWEKMLQDTEKRR-EKKRKKOKNLDNRHEPEK 192
Db 424 LTIECVSQIVLHKNGTDPDFCKRHLOIDIERLVDOMIDTKVEKSEAKATELEKLDSEL 483
QY 193 VPRAPHRDRRWQKL-----AQGPDLAEDDANL-----LHKHIEVANGPASH 234
Db 484 TAR--HELOVEMKMKENDFEQKLQDLQEKDLDLSEKQITAKQDLAEVSKLTGEVAK 541
QY 235 FETRPQTYVDHMDGYSLSALPFSQMSSELLTRA-----EERVLVRPHEPPPPPPMHGAGD 289
Db 542 LSKELEDARNEMASLSAVVAVSVSSAAVPPAPPLPGSGIVIPP-PPPPPLPG--- 596
QY 290 AKPIPTCISSAVGLIENRQSPATGRTVPVSPTP-----PPPPPLPSALSTSSLRASMT 345
Db 597 -----GVV---PESP-----PLPPTCIPPPPPPLPGG-----A 621
QY 346 STPPPP-----VPPPPPPATALQAPAVPPPPAPLQIAPGLVHPAPPIAPPLVQSP 398
Db 622 CIPPPPLPGSAAIPPPPLPGVA---SIPPPP-PLPGATAI---PPPPPLPGATAIPP 674
QY 399 PVARAAPVCETVPVHPLPQGEVGLPPPPPPPP-----LPPGIRPSSPVTVTALAHPPSG 454
Db 675 P-----PLPG--TGIPPPPPPLPGSVGVPPPPPLPGGP-----G 707
QY 455 LHPTSTAPGPHVPLMPPSPPSQVIPASEP-----KRHPSTLPVISDARSVLEAIRKGIQ 510
Db 708 LPPPPPPPPG--APGIPPPPPPGMGVPPPPPPFGVGVAAPVLPGLTPKKV-----YKPEVQ 761
QY 511 LRK-----VEQREOE-----AKHERIEND 530
Db 762 LRRPNMSKFEADLSQDCFQWTKVKEDRFENN 792
RESULT 4
DIAL_HUMAN
ID DIAL_HUMAN STANDARD; PRT; 1248 AA.
AC O06010;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 1 (DIAPHANOUS-RELATED FORMIN 1) (DRFL).
GN DIAPH1 OR DIAP1.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98028756; PubMed=9360932;
 RA Lynch E.D., Lee M.K., Morrow J.E., Welch P.L., Leon P.E., King M.-C.;
 RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
 RT homolog of the Drosophila gene diaphanous.";
 RL Science 278:1315-1318(1997).
 RN [2]
 RP SEQUENCE OF 218-817 FROM N.A.
 RC TISSUE=Ovarian carcinoma;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Watanabe S., Hoshino T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NED0 human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
 CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
 CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
 CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
 CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
 CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
 CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
 CC ACTIN POLYMERIZATION IN HAIR CELLS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
 CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.
 CC -1- DOMAIN: DFRS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
 CC RHO-GTP ACTIVATES THE DFRS BY DISRUPTING THE GBD-DAD INTERACTION
 CC (BY SIMILARITY).
 CC -1- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE AUTOSOMAL DOMINANT
 CC NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).
 CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
 CC SUBFAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
 CC TO INCLUDE INTRONIC SEQUENCE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AF051782; AAC05373.1;
 DR EMBL; AK023345; BAB14533.1; ALT_SEQ.
 DR MIM; 602121;
 DR MIM; 124900;
 DR InterPro; IPR003104;
 DR Pfam; PF02181; FH2; 1.
 KW Coiled coil; Repeat; Deafness.
 FT DOMAIN 63 260
 FT Coiled coil; Repeat; Deafness.
 FT DOMAIN 157 457
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 460 563
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 588 743
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 748 1190
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 1015 1172
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 1173 1187
 FT Coiled coil (POTENTIAL).
 FT DOMAIN 1189 1192
 FT Coiled coil (POTENTIAL).
 FT ARG/LYS-RICH (BASIC).
 FT CONFLICT 804 804
 FT TSKA (IN REF. 2).
 SQ SEQUENCE 1248 AA; 138978 MW; EDIF5147CFF9A886 CRC64;

Query Match 10.5%; Score 313.5; DB 1; Length 1248;
 Best Local Similarity 25.2%; Pred. No. 3.7e-08;
 Matches 146; Conservative 58; Mismatches 177; Indels 199; Gaps 34;
 QY 51 IFGELFNEAHFSFRVNSLQERY-----DRLSVSVTQDLPKEELS--- 91
 DB 308 LINALITPAEELDFRVHRSLEMLRGLHVLQDLREIENEDMRVOLNVFDEGEEDSYDL 367
 QY 92 ---LQDITMRKAFRSSITQDQQLDPLPIQFTYDVCEQPPPLNLITP---YRDGK 145
 DB 368 KGRLLDIRMDDFNEVEF-----ILLNTVKDSKAEHPFLSLQLHLLLRNDYE 416
 QY 146 EGLAFY-----TNPYSY---FFDLWKEKMLQDTEKRR-EKRRKQKQKMLD 185
 DB 417 ARPQYKLIIEICISQIVLHKNAGADPFKRLHQLQIEGLIDQMDTKVKESEAKAAELE 476
 QY 186 RPHEPEKVPRAHPDRRREWKLAQGPPEAEADANLLKHIEVANGPASHFFTRPTVYVDH 245
 DB 477 KKLDSLTAR--HELQVEMKKWESDFEQKLQDLQ-----GEKDALHSE-KQOIATEK 525
 QY 246 MDGYSLSLSALFQMSSELLTRAEE-----RVLYRPHPEP-----PPPPMHGAGD 289
 DB 526 QDLAEVSQLT-GEVAKLTKELEDAKEMASLSAAAIPTPPSPVPAPPLPG--- 581
 QY 290 AKIPTCISSATGLIENRPSQATGRTPVFSVPTPPPPPLPSALSTSSLRASMTSTPP 349
 DB 582 -----DSGTIIIPPPAPGSGSTTPP-PPPPPPPPPLPGGTAIS-----PP 621
 QY 350 PP-----VPPPPP-PPATALQAP-----AVPPPPAPLQIAFVLHPAPPPIAPPL-- 393
 DB 622 PPLSGDATIPPPPLPEGVGPSPSSLPGLGTAIPPPP-PL---PGSARIPPPP--PPLPG 675
 QY 394 ---VQSPVPAARAPVCTVPHLPQGEVQGLPPPPPP-----PPLPPGIRPSSPVTVT 445
 DB 676 SAGIPPPPP-----PLP-GEA-GMPPPPPPLPGFGIPIPPPPPPGPG--- 715
 QY 447 ALAHPPCSLHPTSTAPGPHVLMPP-----SPSQVIP-ASEPKRHPSTLPSIDARSVL 501
 DB 716 -----GIPPPP--PGMGMPPPPPPGFGVPAAPVLPFGLTGPKK----- 750
 QY 502 LEAIRKGIQLRK-----VEQREQE-----AKHERIEND 530
 DB 751 --LYKPEVQLRRRNKSKLVAEDLSQDCFWTKVKEDRFENN 788
 RESULT 5
 EXTN_TOBAC STANDARD: PRT; 620 AA.
 ID AC EXTN_TOBAC
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRGPNT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 RT glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.

CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: X13885; CAA32090.1; -
DR PIR; S06733; S06733.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 ?
FT CHAIN ? 620
FT REPEAT 70 73 EXTENSIN.
FT REPEAT 148 151 H-A-P-P.
FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT REPEAT 229 235 1.
FT REPEAT 236 242 2.
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 10.4%; Score 309; DB 1; Length 620;
Best Local Similarity 34.7%; Pred. No. 3e-08;
Matches 90; Conservative 9; Mismatches 90; Indels 70; Gaps 15;

QY 279 PPPPMHAGDAKPIPTCISATGLIENRPSQATGRTVPFVSPT-----pppp-- 327
DB 370 PPPPP-----SSPPPSPPPTTVEQSPPPPPAYSPPLPAPPTYSPPPTTSPPTTY 423
QY 328 --PPPLPSALSTSLRASMTSTPPPV-----PPPP--PPATALQAP---AVPPPPAPLQ 376
DB 424 AQPPLPP-----TYSPPPPAYSPPPPTTSPPTTSPPPPPAYAQPPPPPTTY 472
QY 377 IAPGLVHPAPPPI-----APPLVQSPPPVARAAP---VCETVPVH--PLPQGEVQGLPPP 426
DB 473 SPPPPAYSPPPSPPTTSPPPPPVQPLPPTTSPPPPPRIHLPPPPHQRPPPTTYGQPPS 532
QY 427 P-----PPPP-----PPPGIRPSPVTV-----TALAHPSPGLH--PTPSTAPGPHV 467
DB 533 PPTTSPPPPPHSPPPHWPHTPTTYQPPSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTT 592
QY 468 PL--MPSPSPQVPASEP 484
DB 593 PTYQPPSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTTSPPTT 611

RESULT 6
WASP_MOUSE STANDARD: PRT: 520 AA.
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE WISKOTT-ALDRICH SYNDROME PROTEIN HOMOLOG (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)

gene is highly conserved and maps near the scurfy (sf) mutation on
the X chromosome.";
Genomics 29:471-477(1995).
-!- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION
(BY SIMILARITY).
-!- DOMAIN: THE WHI (WASP HOMOLOGY 1) DOMAIN MAY BIND A PRO-RICH
LIGAND.
-!- SIMILARITY: CONTAINS 1 GBD DOMAIN.
-!- SIMILARITY: CONTAINS 1 WHI DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U54788; AAC52556.1; -
DR MGD; MGI:105059; Wasp.
DR InterPro; IPR000095; -
DR InterPro; IPR001960; -
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00568; WHI; 1.
DR PROSITE; PS50108; GBD; 1.
KW Repeat.
FT DOMAIN 41 147 WHI.
FT DOMAIN 240 259 GBD.
FT REPEAT 354 363 GRSGPLPPXP MOTIF 1.
FT REPEAT 393 402 GRSGPLPPXP MOTIF 2.
FT DOMAIN 162 167 POLY-PRO.
FT DOMAIN 314 321 POLY-PRO.
FT DOMAIN 324 341 POLY-GLY.
FT DOMAIN 368 373 POLY-PRO.
FT DOMAIN 376 379 POLY-PRO.
FT DOMAIN 397 403 POLY-PRO.
FT DOMAIN 408 424 POLY-PRO.
FT DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match 10.0%; Score 297; DB 1; Length 520;
Best Local Similarity 24.3%; Pred. No. 9.4e-08;
Matches 133; Conservative 36; Mismatches 138; Indels 240; Gaps 27;

QY 122 QETDYDCEQPPPLNLTPT-----RDKGKGLKF--YTNPSTFDLWKEKMLQDTEK 174
DB 101 QELYSQLVLTPT-----TPFFHTFAGDDCQGLNFADESEAAFRALVQEKI-----OKRN 151
QY 175 EKRKOKKNDLRPHPEKVP-----
DB 152 QROSGERRQLPPPPAPLINEERRGGLPPVPHPGDHGGSGGSLGLVTVDIQNDITS 211
QY 196 -----APHDRRRWQKLAQPELAEDDANLLHKHIEVANGPASHFETRPQTYVD 244
DB 212 SRYRGLPAPCGPTDKKRSCKKISRADICAPSG---FKHV-----SHVWDNQDGF 261
QY 245 HMDGSYLSALPFSQ--MSEL-LTRAEERVLV-----RPHE--ppppp 282
DB 262 VNNLDPLRLS-LFSAIGISEAQLTDAETSKLIYDFIEDQGLEAVRQEMRRQEPPLPPPP 320
QY 283 PMH-----GAGDAKPIPTCISATGLIENRPSQATGRTVPFVSPTTTP----- 325
DB 321 PCRGGGGGGGGGGGGGGGGQPL-----RP--PVVGSNKGSGGLPPVPMG 365
QY 326 -PPPPPLPSALSTSLRASMTSTPPPP-----VPPPPPPATALQAPVPPPPAPLQIAPGV 381
DB 366 GAPPPPTTPTG-----PPPPGGRGPPPPPPPPATGTRSGP--PPPLPGAGGP-- 408
QY 382 LHPAPPPIAPPLVQSPPPVARAAPVCTVPHPLPQGEVQGLPPPPPPPPPPPPPPPPSS 441
|||||

RX MEDLINE=96133286; PubMed=8528199;
 RA Derry J.M.J., Kerns J.A., Weinberg K.I., Ochs H.D., Volpini V.,
 RA Estévil X., Walker A.P., Francke U.;
 RT "WASP gene mutations in Wiskott-Aldrich syndrome and X-linked
 RT thrombocytopenia.";
 RL Hum. Mol. Genet. 4:1127-1135(1995).
 RN [7]
 RP VARIANTS WAS V-56 AND E-236.
 RX MEDLINE=95315993; PubMed=7795648;
 RA Villa A., Notarangelo L., Macchi P., Mantuano E., Cavagni G.,
 RA Brugnani D., Strina D., Patrosso M.C., Ramenghi U., Sacco M.G.,
 RA Ugazio A., Vezzoni P.;
 RT "X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic
 RT diseases with mutations in the WASP gene.";
 RL Nat. Genet. 9:414-417(1995).
 RN [8]
 RP VARIANTS WAS TRP-43; MET-45; MET-75 AND CYS-86.
 RX MEDLINE=97272113; PubMed=9126958;
 RA Remold-O'Donnell E., Cooley J., Scherbin A., Hagemann T.L.,
 RA Kwan S.-P., Kenney D.M., Rosen F.S.;
 RT "Variable expression of WASP in B cell lines of Wiskott-Aldrich
 RT syndrome patients.";
 RL J. Immunol. 158:4021-4025(1997).
 RN [9]
 RP VARIANTS WAS LYS-31 AND MET-45.
 RX MEDLINE=97253413; PubMed=9098856;
 RA Ariga T., Yamada M., Sakiyama Y.;
 RT "Mutation analysis of five Japanese families with Wiskott-Aldrich
 RT syndrome and determination of the family members' carrier status
 RT using three different methods.";
 RL Pediatr. Res. 41:535-540(1997).
 RN [10]
 RP VARIANTS WAS MET-75; LEU-84; ASP-89 AND LYS-133.
 RX MEDLINE=98350091; PubMed=9683546;
 RA MacCarthy-Morrogh L., Gaspar H.B., Wang Y.-C., Katz F., Thompson L.,
 RA Layton M., Jones A.M., Kinnon C.;
 RT "Absence of expression of the Wiskott-Aldrich syndrome protein in
 RT peripheral blood cells of Wiskott-Aldrich syndrome patients.";
 RL Clin. Immunol. Immunopathol. 88:22-27(1998).
 RN [11]
 RP VARIANTS WAS VAL-56.
 RX MEDLINE=98378988; PubMed=9713366;
 RA Facchetti F., Blanzuoli L., Vermi W., Notarangelo L.D., Gillani S.,
 RA Fiorini M., East A., Stewart D.M., Nelson D.L.;
 RT "Defective actin polymerization in EBV-transformed B-cell lines from
 RT patients with the Wiskott-Aldrich syndrome.";
 RL J. Pathol. 185:99-107(1998).
 RN [12]
 RP VARIANTS WAS LYS-133.
 RX MEDLINE=98092074; PubMed=9445409;
 RA Parolini O., Rasmann G., Haas O.A., Pawlowsky J., Gadner H.,
 RA Knapp W., Holter W.;
 RT "X-linked Wiskott-Aldrich syndrome in a girl.";
 RL New Engl. J. Med. 338:291-295(1998).
 RN [13]
 RP VARIANTS WAS ARG-73; MET-75; CYS-83; CYS-86 AND LYS-133.
 RX MEDLINE=99374925; PubMed=10447259;
 RA Lemahieu V., Gastier J.M., Francke U.;
 RT "Novel mutations in the Wiskott-Aldrich syndrome protein gene and
 RT their effects on transcriptional, translational, and clinical
 RT phenotypes.";
 RL Hum. Mutat. 14:54-66(1999).
 CC -!- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
 CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION.
 CC -!- SUBUNIT: BINDS TO CDC42, RAC, NCK, FYN, SRC KINASE FGR, BTK, ABL,
 CC WIP, AND TO THE P85 SUBUNIT OF PLC-GAMMA.
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE THYMUS. ALSO
 CC FOUND, TO A MUCH LESSER EXTENT, IN THE SPLEEN.
 CC -!- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
 CC LIGAND.
 CC -!- DISEASE: DEFECTS IN WASP ARE THE CAUSE OF WISKOTT-ALDRICH SYNDROME
 CC (WAS), AN X-LINKED RECESSIVE IMMUNODEFICIENCY CHARACTERIZED BY
 CC ECZEMA, THROMBOCYTOPENIA, RECURRENT INFECTIONS, AND BLOODY

CC DIARRHEA. DEATH USUALLY OCCURS BEFORE AGE 10.
 CC -!- DISEASE: DEFECTS IN WASP ARE THE CAUSE OF ISOLATED X-LINKED
 CC THROMBOCYTOPENIA (XLT). XLT IS CLINICALLY MILD WITH SMALL
 CC PLATELETS AND SUBCLINICAL LEUCOCYTE ABNORMALITIES.
 CC -!- SIMILARITY: CONTAINS 1 GBD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U12707; AAA62663.1; ALT_SEQ.
 CC EMBL; U18935; AAA60381.1; -.
 CC EMBL; U19927; AAC50140.1; -.
 CC EMBL; AF196970; AAF06804.1; -.
 CC MIM; 301000; -.
 CC InterPro; IPR0000095; -.
 CC InterPro; IPR001960; -.
 CC Pfam; PF00786; PBD; 1.
 CC Pfam; PF00568; WH1; 1.
 CC PROSITE; PS50108; GBD; 1.
 CC Repeat; Disease mutation.
 CC WH1.
 CC GBD.
 CC GRSGPLPPXP MOTIF 1.
 CC GRSGPLPPXP MOTIF 2.
 CC POLY-PRO.
 CC POLY-PRO.
 CC POLY-PRO.
 CC POLY-PRO.
 CC POLY-PRO.
 CC POLY-PRO.
 CC ASP/GLU-RICH (ACIDIC).
 CC L -> F (IN THROMBOCYTOPENIA).
 CC /FTIG-VAR_005823.
 CC MISSING (IN THROMBOCYTOPENIA).
 CC /FTIG-VAR_005824.
 CC E -> K (IN WAS).
 CC /FTIG-VAR_005825.
 CC C -> W (IN WAS; MODERATE FORM).
 CC /FTIG-VAR_008105.
 CC T -> M (IN WAS).
 CC /FTIG-VAR_008106.
 CC T -> I (IN THROMBOCYTOPENIA).
 CC /FTIG-VAR_005826.
 CC A -> V (IN THROMBOCYTOPENIA).
 CC /FTIG-VAR_005827.
 CC C -> R (IN WAS; SEVERE FORM).
 CC /FTIG-VAR_008107.
 CC V -> M (IN THROMBOCYTOPENIA).
 CC /FTIG-VAR_005828.
 CC S -> P (IN ATTENUATED WAS).
 CC /FTIG-VAR_005829.
 CC Y -> C (IN XLT).
 CC /FTIG-VAR_008108.
 CC F -> L (IN WAS; SEVERE FORM).
 CC /FTIG-VAR_008109.
 CC R -> C (IN WAS).
 CC /FTIG-VAR_005832.
 CC R -> H (IN WAS).
 CC /FTIG-VAR_005830.
 CC R -> L (IN WAS).
 CC /FTIG-VAR_005831.
 CC G -> D (IN WAS; MILD FORM).
 CC /FTIG-VAR_008110.
 CC W -> C (IN ATTENUATED WAS).
 CC /FTIG-VAR_005833.
 CC E -> K (IN WAS).
 CC 131 131

```
FT VARIANT 133 133 /FTID=VAR_005834.
FT E -> K (IN WAS: SEVERE FORM).
FT /FTID=VAR_005835.
FT VARIANT 187 187 G -> C (IN WAS).
FT /FTID=VAR_005836.
FT VARIANT 236 236 A -> E (IN THROMBOCYTOPENIA).
FT /FTID=VAR_005837.

Query Match 9.4%; Score 280; DB 1; Length 502;
Best Local Similarity 23.6%; Pred. No. 5.7e-07;
Matches 126; Conservative 37; Mismatches 142; Indels 228; Gaps 25;

Qy 122 QETDVCEQPPPLNLTPT-----RDGKEGLKF--YTNPSYFFDLNKEKMLQDTEKRR 174
Db 99 QELYSQVLYSTP-----TPFFHTFAGDCOAGLNFADAQAQFALVQEKI-----QKRN 149
Qy 175 EKRKQKQKNDLRPEKVPK-----
Db 150 QKQSGDRQLPPTPTANEERRGGLPLPLHPGGDQGGPPVGPLSLGLATVDIQNPDIITS 209
Qy 196 -----APHDRREWKLAQGPDLAEDDANLLKHIEVANGPASHEFETRPQTVVD 244
Db 210 SRYRGLPAPGSPADKKRSGKKKTSKADIGAPSG---FKHV-----SHVGNPDQNGFD 259
Qy 245 HMDGSYSLSALPFSQ--NSEL-LTRAERVLV-----RPHEP-PPPPP 283
Db 260 VNNLDPLRLS-LFSRAGISEAQLTDAETSKLIYDFIEDQGLEAVROEMRQEPPLPPPP 318
Qy 284 MHGADAKPIPTCISATGLIENRQSPATGTPVFSVPTPP-----PPPPPLPSALSTS 338
Db 319 PSRGNQLPRPIVGG-----NKGRS-----GPLPPVPLGIAPPPPTPRG----- 358
Qy 339 SLRASMTSTPPPP---VPPPPPPATALQAPVPPPPAPLQIAPGLHAPPPPIAPPLVQ 395
Db 359 -----PPPPGRRGPPPPPPPPGATGSGPLPPPPPG----- 387
Qy 396 PSPVARAAPVCETVPVHPLPQGEVQGLPPPPPPPPPPPPPPGIRPSSPVTVTALAHPSSGL 455
Db 388 -----AGPPPPPPPPPPPP-----PSSGN 408
Qy 456 HPTSTAPGHVLMPPSPSSQVTPASEPKRHPSTPLVSDARSVLLAIRKGQOLRKE 515
Db 409 GPAP-----PPLPPA-LVPAG-----GLAP--GGGRGALLDQIRQGIQLNKP 448
Qy 516 EQREQEA-----KHRENDVATILSRRIAYEYSDSE-----DDSEFDEVD 556
Db 449 GAPESSALQPPQSSGLVGLMHVMQKRRAIHSDEGEQAGDEDEDD 501

RESULT 12
ID EBN2_EBV STANDARD; PRT; 487 AA.
AC P12978;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE EBNA-2 NUCLEAR PROTEIN.
GN BYRF1.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
RL Nature 310:207-211(1984).
RN [2]
RP SUBCELLULAR LOCATION, AND PHOSPHORYLATION.
RX MEDLINE=90266473; PubMed=2161150;
```

```
RA Petti L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
RN [3]
RP DOMAINS.
RX MEDLINE=91202599; PubMed=1850028;
RA Cohen J.I., Wang F., Kieff E.;
RT "Epstein-Barr virus nuclear protein 2 mutations define essential
RT domains for transformation and transactivation.";
RL J. Virol. 63:2545-2554(1991).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. TRANSACTIVATES THE EXPRESSION
CC OF LMP-1.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC -1- PTM: PHOSPHORYLATED.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01555; CAA24877.1; ALT_INIT.
CC TRANSPAC; T01618;
CC Transcription regulation; Activator; Nuclear protein; DNA-binding;
CC Phosphorylation; Repeat.
CC POLY-PRO.
CC DOMAIN 59 100 6 X 2 AA TANDEM REPEATS OF R-G.
CC DOMAIN 345 356
CC SEQUENCE 487 AA; 52544 MW; DEF40D7F8ED61DIA CRC64;

Query Match 9.28; Score 275.5; DB 1; Length 487;
Best Local Similarity 27.1%; Pred. No. 9e-07;
Matches 96; Conservative 24; Mismatches 73; Indels 167; Gaps 18;

Qy 240 QTYVDHMD---GYSLSALPFSQMSSELLTRAE-----ERVLPHEPPPPPPMHG 286
Db 12 QTYHLIVDTDSLGNPSLSVIPSNPYEQSLDTPLPLTIFVGENTGVPPPLPPPPPP--- 68
Qy 287 AGDAKPIPTCISATGLIENRQSPATGTPVFSVPTPPPPPPPLPSALSTSLSRASMTS 346
Db 69 -----PPPP-----PPPPPPPPPP----- 83
Qy 347 TPPPPVPPPPPPATALQAPVPPP-----PAPLIQIAP----- 379
Db 84 PPPPPSPPPPPPPPP-----PPQRRDAWTQBPSPDLRDLGYDVGHGLASAMRLM 135
Qy 380 -----GVLHPAPPPIAPP--LVQSPPPVARAAPVCETVPVHPLPQGEVQ-- 422
Db 136 MANYIVQRSRGRGLILPQGTAPQARLVQPHVPLR--PTAPT-LSPLSQPLTPPQ 192
Qy 423 ---Lpp-PPPPPPPLPPPGI-----RPSSEVT-----VTALAHPPSGLHPTPTAPGPHVPL 469
Db 193 PLMPPPRPTPTPLPPATLTVPPRPTRTTLPPTPLTLVLRTELQPTPS-PPRMHLPV 251
Qy 470 M-----PPSPSOVIPASEPKRHSHTLPVI 494
Db 252 LHVPDQSMHPLTHQSTNDPDSPEFRSPTVFYINPPMPLPPSQLPPPAAPQPPP--GVI 309
Qy 495 SD 496
Db 310 ND 311

RESULT 13
ID CAPU_DROME STANDARD; PRT; 1059 AA.
AC Q24120; Q9VQV8;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
```



```
FT * DOMAIN 870 970 FH1 (PRO-RICH).
FT DOMAIN 985 1426 FH2.
FT DOMAIN 1305 1378 COILED COIL (POTENTIAL).
FT DOMAIN 198 203 POLY-SER.
FT DOMAIN 861 864 POLY-SER.
FT DOMAIN 885 892 POLY-PRO.
FT DOMAIN 911 925 POLY-PRO.
FT DOMAIN 929 940 POLY-PRO.
FT DOMAIN 951 962 POLY-PRO.
FT DOMAIN 966 970 POLY-PRO.
FT DOMAIN 977 981 POLY-SER.
FT VARSPLIC 1252 1287 MISSING (IN ISOFORM IB).
FT VARSPLIC 625 722 MISSING (IN ISOFORM II).
FT VARSPLIC 626 627 IA -> SV (IN ISOFORM III).
FT VARSPLIC 628 1468 MISSING (IN ISOFORM III).
SQ SEQUENCE 1468 AA; 163809 MW; EF2FB1E9CA9DAF43 CRC64;

Query Match 9.1%; Score 272.5; DB 1; Length 1468;
Best Local Similarity 23.0%; Pred. No. 3.6e-06;
Matches 121; Conservative 50; Mismatches 173; Indels 181; Gaps 20;

QY 53 GE-LFNEAHSFVRNLSQERVRLSVTLDPKKEELSLODITMRKAFRSSTIQDQQL 111
Db 549 GEVFNVEYPSRKNDVSSGPPSADTL-----EPSSTTKVTETKGASPTSLRASQ- 596

QY 112 FDRKTLPIQLQETDYDCEQPP-----PLNLTPTYRDGKEGLKFTYNPSYFFDLWKEKML 166
Db 597 -----TWLVSEASEAKGLGPEKITAPPQHQLPGI---ASEGFFCDNFKEQTA 641

QY 167 QDTEDK-----RKEKQKQKQKNDLRPHPEKEKVPRAHDR-----RR 202
Db 642 KDLFNKDGWVWFGYRAGPCPFLLHHEEKEKTSRSELYLDLNFQDQSPTEQDDRTPGRLQA 701

QY 203 EW-----QKLAQGPDLAEDANLLH---KHIEVANGPASHFETRPOTYVDHMDGSY 250
Db 702 VWPPPKTKDTEKVGKLYTEAYEQAILHLKREHKEEITLQAFELK--TF--HNGEH 757

QY 251 SLSALPFSQMSLELLTRAER-----VLVRPHEPPPPPPMHG----- 286
Db 758 ALVTARLEEAIEIENLKQLEKREGECEMRDVCISTDDCSKAFRNVCIQTDRETLKPC 817

QY 287 -----AGDAKPICTCISSATGLIENRPSQSPATGRTPFV 320
Db 818 DAESKATRRSQIVPKKLTISLTQLSPSKOSKDIHAPFQREGTSSSSQK-----I 868

QY 321 SPTPPPPPPPLPSALSTSSLRASMTSTPPPPVPP-----PPPPATALQAPAVPPPP 372
Db 869 SPAPPTPPPLPPL-----IPPPPLPGLGLPAPPPIPPVCVSPPPPPPP 918

QY 373 APLQIAPGVLHPAPPPIAPPLVQFSPVARAAVCETVPHPLPQGEVQGLPPPPPPPL 432
Db 919 PPTVPVPSDGGPPPPPPPPPL-----PNVLALP-----NSGGPPPPPPPPPP 959

QY 433 PPGCIRPSSPVVYALAHPPSLH--PTPSTAGPHVPLMPPSP 475
Db 960 PPGGLAPP-----PPGLSFLSSSSQYPRKPAIEPSCP 994
```

Search completed: June 20, 2001, 12:10:54
Job time: 194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:11:41 ; Search time 13.03 Seconds
(without alignments)
1469.595 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 559
Sequence: 1 MPLVKRNDPRHLCFTALPR.....AVEYSDSDSEDFEVDWLE 559

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 93435 seqs, 34255486 residues
Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries
Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	559	100.0	559	1 Y269_HUMAN	Q92558 homo sapien
2	12	2.1	281	1 FASL_HUMAN	P48023 homo sapien
3	12	2.1	389	1 NDRP_MOUSE	Q03173 mus musculus
4	11	2.0	314	1 Y009_HUMAN	Q15390 homo sapien
5	11	2.0	477	1 CAP2_HUMAN	P40123 homo sapien
6	11	2.0	971	1 Y029_HUMAN	Q15032 homo sapien
7	11	2.0	994	1 CLC1_RAT	P35524 rattus norv
8	10	1.8	185	1 NO20_SOYBN	P08960 glycine max
9	10	1.8	211	1 YN03_CAEEL	Q03607 caenorhabdi
10	10	1.8	288	1 SMN_MOUSE	P97801 mus musculus
11	10	1.8	367	1 P53_CHICK	P10360 gallus gall
12	10	1.8	477	1 CAP2_RAT	P52481 rattus norv
13	10	1.8	551	1 CAP_SCHPO	P36621 schizosacch
14	10	1.8	809	1 CN4D_HUMAN	Q08499 homo sapien
15	10	1.8	853	1 DIA3_HUMAN	Q9nsv4 homo sapien
16	10	1.8	1108	1 CN3B_RAT	Q63085 rattus norv
17	10	1.8	1171	1 DIA3_MOUSE	Q05859 mus musculus
18	10	1.8	1206	1 FM14_MOUSE	O60610 homo sapien
19	10	1.8	1248	1 DIA1_HUMAN	O08808 mus musculus
20	10	1.8	1255	1 DIA1_MOUSE	P40450 saccharomyc
21	10	1.8	1375	1 BNR1_YEAST	Q05860 mus musculus
22	10	1.8	1468	1 FM1_MOUSE	P34433 caenorhabdi
23	9	1.6	244	1 YL53_CAEEL	Q10390 mycobacteri
24	9	1.6	299	1 MMS3_MYCTU	P22421 human papil
25	9	1.6	304	1 VE4_HPV47	P52952 homo sapien
26	9	1.6	324	1 HK25_HUMAN	P53339 homo sapien
27	9	1.6	338	1 FOSB_HUMAN	PI3346 mus musculus
28	9	1.6	338	1 FOSB_MOUSE	Q91907 xenopus lae
29	9	1.6	340	1 GBX2_XENLA	P52951 homo sapien
30	9	1.6	348	1 GBX2_HUMAN	P48031 mus musculus
31	9	1.6	348	1 GBX2_MOUSE	P14652 homo sapien
32	9	1.6	356	1 HXB2_HUMAN	O24175 oryza sativ
33	9	1.6	389	1 FL_ORYSA	

34	9	1.6	412	1 ALF_PETHY	Q22621 petunia hybr
35	9	1.6	421	1 ACRO_HUMAN	P10323 homo sapien
36	9	1.6	435	1 KICH_MOUSE	O54804 mus musculus
37	9	1.6	440	1 FXGA_CHICK	Q98937 gallus gall
38	9	1.6	440	1 G3PT_MOUSE	Q64467 mus musculus
39	9	1.6	450	1 CYSP_TRYBB	P14658 trypanosoma
40	9	1.6	453	1 KICH_RAT	Q01134 rattus norv
41	9	1.6	487	1 EBN2_EBV	P12978 epstein-bar
42	9	1.6	502	1 WASP_HUMAN	P42768 homo sapien
43	9	1.6	515	1 P2BB_MOUSE	P48453 mus musculus
44	9	1.6	518	1 TPM4_DROME	P49455 drosophila
45	9	1.6	520	1 WASP_MOUSE	P70315 mus musculus

ALIGNMENTS

RESULT 1					
Y269_HUMAN					
ID	Y269_HUMAN	STANDARD;	PRT;	559 AA.	
AC	Q92558;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269.				
GN	KIAA0269.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone marrow;				
RX	MEDLINE=97919144; PubMed=9039502;				
RA	Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,				
RA	Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;				
RT	"Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280), deduced by analysis of cDNA clones from cell line KG-1 and brain.";				
RT	DNA Res. 3:321-329(1996).				
CC	-1- SIMILARITY: TO C-ELEGANS R06C1.B.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; D87459; BAA13399.1;				
KW	Hypothetical protein.				
FT	DOMAIN 278 283 POLY-PRO.				
FT	DOMAIN 322 332 POLY-PRO.				
FT	DOMAIN 348 359 POLY-PRO.				
FT	DOMAIN 369 374 POLY-PRO.				
FT	DOMAIN 424 435 POLY-PRO.				
SQ	SEQUENCE 559 AA; 61652 MW; 44B4527BDB77BC6E CRC64;				

Query Match	100.0%;	Score 559;	DB 1;	Length 559;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 559;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MPLVKRNDPRHLCFTALPRGKNELECVTNISLANIROLSSLSKYAEDIFGELFNEAH	60	
Db	1	MPLVKRNDPRHLCFTALPRGKNELECVTNISLANIROLSSLSKYAEDIFGELFNEAH	60	
QY	61	SFSFRVNSLQERYDRLSVSVTQLDPKKEELSLODITMRKAFRSTTODQQLFDRKTLPIP	120	
Db	61	SFSFRVNSLQERYDRLSVSVTQLDPKKEELSLODITMRKAFRSTTODQQLFDRKTLPIP	120	
QY	121	LQETYDVCEQPPPLNLTTPYRDDGKGLKFYTPNSYFFDLWKMKMLQDTEKRRKKQK	180	

[illegible]

RESULT	2	FASL_HUMAN	STANDARD;	PRT;	281 AA.
AC	1	FASL_HUMAN	STANDARD;	PRT;	281 AA.
AD	2	P48023;			
AT	3	01-FEB-1996 (Rel. 33, Created)			
DT	4	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	5	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	6	FAS ANTIGEN LIGAND (APOPTOSIS ANTIGEN LIGAND) (APTL).			
OS	7	FNFSF6 OR APTL1G1 OR FASL.			
GN	8	Homo sapiens (Human).			
OC	9	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	10	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	11	NCBI_Taxid=9606;			
RP	12	SEQUENCE FROM N.A.			
RP	13	SEQUENCE FROM N.A.			
RR	14	MEDLINE=95105731; PubMed=7528780;			
RR	15	Algeron M.;			
RT	16	"Fas ligand mediates activation-induced cell death in human T			
RT	17	lymphocytes.";			
RL	18	J. Exp. Med. 181:71-77(1995).			
RP	19	(2)			
RP	20	SEQUENCE FROM N.A.			
RR	21	MEDLINE=95127560; PubMed=7826947;			
RR	22	Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;			
RT	23	"Human Fas ligand: gene structure, chromosomal location and species			
RT	24	specificity.";			
RL	25	Int. Immunol. 6:1567-1574(1994).			
RP	26	(3)			
RP	27	SEQUENCE FROM N.A.			
RR	28	Schaetzlein C.E., Poehlimann R., Philippson P., Eibel H.;			
RR	29	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
RP	30	(4)			
RP	31	SEQUENCE FROM N.A.			
RR	32	MEDLINE=95071350; PubMed=7980502;			
RR	33	Mita E., Hayashi N., Iio S., Takehara T., Hijioaka T., Kasahara A.,			
RR	34	Fusamoto H., Kanada T.;			
RT	35	"Role of Fas ligand in apoptosis induced by hepatitis C virus			
RT	36	infection.";			
RL	37	Biochem. Biophys. Res. Commun. 204:468-474(1994).			
RP	38	(5)			
RP	39	SEQUENCE FROM N.A.			

Wilkinson J.;
Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE OF 1-10 FROM N.A.
TISSUE=Blood;
Matsumura M., Nakanishi Y., Ohba Y.;
Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT
TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN
CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.
FAS-ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF
PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE
T CELLS, OR BOTH.
-!- SUBUNIT: HOMOTRIMER (PROBABLE).
-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED
INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FORM THE CELL
SURFACE.
-!- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X89102; CAA61474.1; -;
EMBL; U08137; AAC50071.1; -;
EMBL; U11821; AAC50124.1; -;
EMBL; D38122; BAA07320.1; -;
EMBL; Z96050; CAB09424.1; -;
EMBL; AB013303; BAA32542.1; -;
HSP; P01375; 2TUN.
MIM; 134638; -;
InterPro; IPR000478; -;
Pfam; PF00229; TNF; 1.
PROSITE; PS00251; TNF_1; 1.
PROSITE; PS50049; TNF_2; 1.
Cytokine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis.
DOMAIN 1 80
TRANSMEM 81 102
DOMAIN 103 281
DOMAIN 4 70
DOMAIN 45 65
DOMAIN 202 233
DISULFD 184 184
CARBOHYD 184 184
CARBOHYD 250 250
CARBOHYD 260 260
SEQUENCE 281 AA; 31485 MW; A8A6EB358246E9BB CRC64;
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
EXTRACELLULAR (POTENTIAL).
PRO-RICH.
POLY-PRO.
BY SIMILARITY
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 281 AA; 31485 MW; A8A6EB358246E9BB CRC64;
Query Match 2.1%; Score 12; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 424 PPPPPPPPLPPP 435
DBB 46 PPPPPPPPLPPP 57

RESULT 3
NDPP_MOUSE
IID NDPP_MOUSE STANDARD; PRT; 389 AA.
Q03173;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
NPC DERIVED PROLINE RICH PROTEIN 1 (NDPP-1).
NDPP1.
Mus musculus (Mouse).
OS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```

Db 120 VLETYNCTDPPPLNLTYPYRDDGTALAFYDPSYFDLWKKMLQDTRDIMKRRKHR 179
QY 181 QKMLDRPHEPEKVPNPRAPHDRRREWOKLAQGPDLAEDDANLLHKHIEVANGPASHFETRPQ 240
Db 180 KKKDNPNRGNVNPRIKTRKEWEKIKMGQEFVESKEKL-----GPGFY----PP 226
QY 241 TYVDHMDGYSLSALPFSOMSELLTRAERVLVRPHEPPPPPPMHGAGDAKPIPTCISSA 300
Db 227 TLV-YONGSIG-----CVENVDASSY-----PPSQSDSASSPSFSE-- 264
QY 301 TGLIENRQSPATGRTPV-----FVSETPPPPPPLPSALSSTSLRASMT 345
Db 265 -----DNLPPPPAEFSYVDNQSGLAGPKRSSVSPSHPPAPPPLGSPSSKPGFAPPP 320
QY 346 STPPPPVPP---PPPPATALQAPVPPPPAPLQIAPGVLP---APPPIAPPLVQPSPP 399
Db 321 APPPPPMIGIPPPPPICGSGCTPPPPSSFP---HPDFAAPPPLPPPPAADYP- 376
QY 400 VARAAVCETVPVHPLPQGEVQGLPPPPPPPLPPGIRPSSPVTVTALAHPPSGLHPTP 459
Db 377 -----TLPPPLSQ-PTRGAPPPPPPPGPP-----PPP 405
QY 460 ST-APGPHVPLMPPSPPSQVIPASEPKRHPSTLPVISDARSVLEAIRKGIQIRKVEEQR 518
Db 406 FTGADGO--PAVPP-----PLSDTTKPKSSLPAISDAHSDLLSAICQGFQLRVVEEQ 456
QY 519 EQAKHERIENDVATILSRRIAVEYSDSEDD-SEFDEVDW 557
Db 457 EQE-KWDVGVNDVATILSRRIAVEYSDSEDDSEFDDGDW 495

RESULT 4
QYUDY7 PRELIMINARY; PRT; 496 AA.
AC QYUDY7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE SCAR2 (FRAGMENT).
GN SCAR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98402540; PubMed=9732292;
RA Bear J.E., Kawai J.F., Saxe C.L. III.;
RT "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT defects in late Dictyostelium development.";
RL J. Cell Biol. 142:1325-1335(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kachesky L.M., Insall R.H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF134304; AAD33053.2;
DR InterPro; IPR002965;
DR InterPro; IPR003124;
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 496 AA; 53963 MW; 8A9D97D0D964734F CRC64;

```

```

Query Match 40.2%; Score 1198; DB 4; Length 496;
Best Local Similarity 47.7%; Pred. No. 4e-74;
Matches 276; Conservative 51; Mismatches 144; Indels 108; Gaps 19;

QY 2 PLVKNRNDPRHLCHTALPRGKNELECVTNISLANIROLSSLSKYAEDIFGELFNEAHS 61
Db 1 PLVTRNIEPHLCRQTLP-SVRSELECVTNITLANVIRQLGSLSKYAEDIFGELFTQANT 59

```

```

QY 62 FSRVNSLOERVDRLSVSVTQDPKKEELSLQDITRKAFRRSTIQDQQLFDRKTIPL 121
Db 60 FASPVSLSAERVDGLQVKVTLQDPKKEEVSQVINTRKAFRRSTIQDQQLFDRNSLPVPV 119
QY 122 QETDYDCEQPPPLNLTYPYRDDGKGLKFTYNPSYFDLWKKMLQDTRDIMKRRKQKQ 181
Db 120 LETYNTCDTPPPPLNLTYPYRDDGTEALKFTYDPSYFDLWKKMLQDTRDIMKRRKHK 179
QY 182 KNLDRPHEPEKVPNPRAPHDRRREWOKLAQGPDLAEDDANLLHKHIEVANGPASHFETRPQ 241
Db 180 EKKDNPNRGNVNPRIKTRKEWEKIKMGQEFVESKEKL-----GPGFY----PPT 226
QY 242 YVDHMDGYSLSALPFSOMSELLTRAERVLVRPHEPPPPPPMHGAGDAKPIPTCISAT 301
Db 227 LV-YONGSIG-----CVENVDASSY-----PPSQSDSASSPSFSE-- 263
QY 302 GLIENRQSPATGRTPV-----FVSETPPPPPPLPSALSSTSLRASMTS 346
Db 264 ---DNLPPPPAEFSYVDNQSGLAGPKRSSVSPSHPPAPPPLGSPSSKPGFAPPPA 320
QY 347 TTPPPVPP---PPPPATALQAPVPPPPAPLQIAPGVLP---APPPIAPPLVQPSPPV 400
Db 321 PPPPPPMIGIPPPPPICGSGCTPPPPSSFP---HPDFAAPPPLPPPPAADYP-- 375
QY 401 ARAAPVCETVPVHPLPQGEVQGLPPPPPPPLPPGIRPSSPVTVTALAHPPSGLHPTPS 460
Db 376 -----TLPPPLSQ-PTRGAPPPPPPPGPP-----PPPF 405
QY 461 T-APGPHVPLMPPSPPSQVIPASEPKRHPSTLPVISDARSVLEAIRKGIQIRKVEEQR 519
Db 406 TGADGO--PAVPP-----PLSDTTKPKSSLPAISDAHSDLLSAICQGFQLRVVEQOE 456
QY 520 QEAKHERIENDVATILSRRIAVEYSDSEDD-SEFDEVDW 557
Db 457 QE-KWDVGVNDVATILSRRIAVEYSDSEDDSEFDDGDW 494

RESULT 5
QYUDY6 PRELIMINARY; PRT; 502 AA.
AC QYUDY6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE SCAR2 (FRAGMENT).
GN SCAR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=99310608; PubMed=10381382;
RA Suetsugu S., Miki H., Takenawa T.;
RT "Identification of two human WAVE/SCAR homologues as general actin
RT regulatory molecules which associate with the Arp2/3 complex.";
RL Biochem Biophys Res Commun. 260:296-302(1999).
DR EMBL; AB026543; BAA81796.1;
DR InterPro; IPR002965;
DR InterPro; IPR003124;
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 502 AA; 55411 MW; 140DD58309345F35 CRC64;

```

```

Query Match 40.1%; Score 1196.5; DB 4; Length 502;
Best Local Similarity 48.5%; Pred. No. 5.1e-74;
Matches 277; Conservative 47; Mismatches 162; Indels 85; Gaps 16;

QY 1 MPLVKNRNDPRHLCHTALPRGKNELECVTNISLANIROLSSLSKYAEDIFGELFNEAH 60

```


George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balowin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P., De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W., Flossler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.

"The genome sequence of *Drosophila melanogaster*," Science 287:2185-2195(2000).

[2] SEQUENCE FROM N.A.

RA Dickson B.J.;

RT "Characterization of the *Drosophila* cytoskeletal regulator SCAR/WAVE";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF003630; AAF53042.1; -;

DR EMBL: AF247763; AAF74194.1; -;

DR FlyBase: FBgn0041781; SCAR.

DR InterPro: IPR003124; -;

DR Pfam: PF02205; WH2; 1.

SQ SEQUENCE 613 AA; 67055 MW; C81578E9FC3F2E95 CRC64;

Query Match 27.1%; Score 807; DB 5; Length 613;
Best Local Similarity 32.3%; Pred. No. 2.4e-47;
Matches 218; Conservative 91; Mismatches 182; Indels 184; Gaps 23;

QY 1 MVLKRNIDPHLCHTALPGIKN--ELECVTNLSIANIIRQLSSLSKYAEDIFGELFNE 58
||| ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 1 MVLPKRSIEPVHVARSVYQDELSQVLEFVTNTLTNIIRQLSSLSKYAEDFVGEALD 60
||| ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 59 AHSFERNVSLQBRVRLSVYVQLDPEKELSLQDTMKRKAFTSSTIDQQLFDRKTLT 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 61 VGNIGRANSLQRIORLAIKVQLDSTVEEVLDTITRKAKFKSAKVFDQOIFSRATMP 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 119 IPLQETVYDCEQPPNLTTPYRDDGKGLKFTNPSYFDFLWKRMLQDTE----DKRK 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 121 APMDVTAQCDKPPPLDKLVYRDDGKGLKFTNPSYFDFLWKRMLKDTERVMDHDKG 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 175 EKPKQKQK-----LDRPHEKVP--RAPHRRREKQK--LAQPELAEDANLLHK--- 223
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 181 KLNRPQDGGAGGAGGAGGNKQKTKIRVPHNTREQQRALVHGVTLMNP--NVYRTPN 238
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 224 ---HIEVANGPASHFEETRP----- 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 239 SMYNEEAGYGNMGVYDTRPPRPNLSIELNRSYQSEQIDGSTYEQLPQMGNOYAAATFGNGM 298
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :

QY 240 ----QTYVDHMD----GSYSLSALPFSQMSSELTRAEE-----ERVLVRPHEPPPPPP 283
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 299 GGPQOHHQHQMYDAGMYQSHAL-YGTGTGQVMSPEIYVGTCTSRKNRPSPQPPAPP 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 284 MHGAGDAKPIPTCISATGLIENRPOSATGRTPVFVSPTPPPPPLPSALTSS----- 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 358 SNGSGGGTGTASNANTPT-----RGRSMSTSRDAL-----PPPPVPDVISPMGNG 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 340 -----LRASMTSTPPPPVPPPPPPPPATALQ-----APAVPP 370
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 406 VNSGHWAALKLGRANSSGAGSPNSVONAMVMTQLSNTFHSIGMTGNOLNSLDLPPP 465
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 371 PPAPLQIAPCVLHPAPPPIAPPLVQSPVPPVAAAPACETVPVHPLPQGEVQGLPPPPPP 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 466 PPVPDQHSP-----KMSPP--NAAP-----PPPPPPPP 490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 431 PLPPPGIRSSPVTVTALAHPPSGLHPTSTADG-----PHVPLMPSPSPQVIPASEPK 485
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 491 PV-EEGMSGNOHTLRPHOILPKSLANGEMQPGQGVPHI-----VAPK 535
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 486 RHPSTLPVISDARSVLLEAIRKGIQLRKVEEQEQAHERIENDVATILSRRIAYEYSD 545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 536 K---MLPPPHDPNDLMLKAIRDGITLRKVEKSEQEKIERNAAPLDVASILARRVAIELSE 592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 546 SED-DSEFDEVDMLE 559
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
DB 593 SEDSDSEDDSEGWNE 607
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
RESULT 8
Q9XVK6 PRELIMINARY; PRT; 507 AA.
AC Q9XVK6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE R06CL.3 PROTEIN.
GN R06CL.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RN SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lighthouse J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R., Smalton N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
RL Nature 368:32-38(1994).
DR EMBL: Z81106; CAB03220.1; -;
DR InterPro: IPR003124; -;
DR Pfam: PF02205; WH2; 1.
DR SMART: SM00246; WH2; 1.
SQ SEQUENCE 507 AA; 55349 MW; C7ABAD18602F590E CRC64;

Query Match 20.1%; Score 598; DB 5; Length 507;

Best Local Similarity 28.8%; Pred. No. 3.5e-33;
Matches 170; Conservative 88; Mismatches 214; Indels 118; Gaps 18;

```

QY 1 MFLVNRNIDPRHLCHTALPRGI-KNELECVTNISLANIIRQLSSLSKYAEDIFGELFNEA 59
D 1 MFLTRAVSPVNLISRGITPSTHRELOCTANGTANLVRQLSSLSKYAENIFGIYHDA 60
QY 60 HSFSPRVNSLQBRVRLSVSVTQDPKEEELSLODITM-KFAFRSSTIQDQOLFDRKTL 119
D 61 MLIHNSKTLOQIRLHKVBDLSNSDQATLNEANRKAFAKSSMLVDQHILDRSTLPT 120
QY 120 PLOEYDYVCEQPPNLTPTDGGKGLKFTNPSYFFDLWKEKMLQDTEKREKR 179
D 121 ALTEIYAKCDPPDLOALNPYRDSIPALSNTNPSFFDLWKEKMLQDTEKREKR 175
QY 180 KOKNLDPRHPEKVPRAHPRDRREMOKLAQGPDLAEDDANLLHKKHIEVANGPASHETRP 239
D 176 -----RVKSPNDGSKSPKRRKQ---PQGP-----LGT 201
QY 240 QTYVDHMGSGYSLSALPFSQMSSELTRAERVLVRPHEPPPPPMHGAADAKPIPTCISS 299
D 202 TAYNDQMQRNROISGRINQNEVSEFPEE-----YQAP-----Q 236
QY 300 ANGLIEN-RPQSPATGRTPVFSPTPPPPPLPSAL---STSSLSRASMTSTPPPPVPPPP 355
D 237 ALGLQNFKNQHPNSNMVAPIGMTMHHHPQONVHPSPQQRGGAPAAARGSPNKRKPTAEAPP 296
QY 356 -----PPPPATALQA-----PAVPPPPAPLQIAPGVH--PAPPPPIAPLVQPSPPVARA 403
D 297 VNLHLPPDPMSILSIDDODDDLPPLPMLMTSIVHQLPAEAPSTIQFVPSAAPTIN 356
QY 404 APVCTETPVV---HPLPGQ---EVQGLPPPPPPPLPPGIRPSSPVTVTA---LAHP 451
D 357 LKICNFCPLCFPHCCPLSASFNSRSLSDPSTTTTNGYKLGCSVICSSGSKGKAG 416
QY 452 PSGLHPT-PSTAGPHVPLMPPSPQVIPASEPKRHPSTPLVDSARSVLLEAIRKGIQ 510
D 417 TAGLYPSADGEAGTNSVPAPKQP-----DARSULLAQIQSGIK 455
QY 511 LRKVEEQREQEAHERIE-NDVATILSRRI-AVEYSDEDDSEFDEVDW 557
D 456 LKKVQRAEAEANAENALNANVAAILKRRMOVGMNDSSSEEGADDEW 505

RESULT 9
Q9XYA8 PRELIMINARY; PRT; 443 AA.
AC Q9XYA8
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SCAR1.
GN SCAR1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Saxe C.L., Bear J.E., Rawls J.F.;
RT "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT defects in late Dictyostelium development.";
RL Cell Biol. Int. 142:1325-1335(1998).
DR EMBL; AF079805; AAD29083.1; -
DR InterPro; IPR002965; -
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 443 AA; 48389 MW; CAA05B0A790454F6 CRC64;

```

Query Match 15.1%; Score 449; DB 5; Length 443;
 Best Local Similarity 26.78; Pred. No. 4.4e-23;
 Matches 150; Conservative 205; Mismatches 205; Indels 132; Gaps 22;

```

QY 1 MFLVNRNIDPRHLCHTALPRGI-KNELECVTNISLANIIRQLSSLSKYAEDIFGELFNEA 59
D 8 LPSVTDNNQP-----ALEGSKQDQIVDTVITSTITVGIINQLTMLVAHNSIFTALANDA 61
QY 60 HSFSPRVNSLQBRVRLSVSVTQDPKEEELSLODITM-KFAFRSSTIQDQOLFDRKTL 117
D 62 NLVTQRIEKLGRIRPLTQSIPIEDYHRTNSIDTMSKPRAEFHADNSERNQHFTHASI 121
QY 118 PLOEYDYVCEQPPNLTPTDGGKGLKFTNPSYFFDLWKEKMLQDTEKREKR 177
D 122 PASINTVTEKCPKPPNQLQDLPYMDGOKSLKLTNPDPFMDWEVVAEQOKLHEEARQKR 181
QY 178 KOKNLDPRHPEKVPRAHPRDRREMOKLAQGPDLAEDDANLLHKKHIEVANGPASHET 237
D 182 ERREARLKKKEKNEV---EVKKVKSVTKVRYDVTGE-----KIINIESPHTS 228
QY 238 RPQTYVDHMGSGYSLSALPFSQMSSELTRAERVLVRPHEPPPPPMHGAADAKPIPTCI 297
D 229 SPQ--IQHOSNN--TATP-----QHTTQHECTNQYQAPPPPL----- 262
QY 298 SSATGLIENRQSPATGRTPVFSPTPPPPPLPSALSTSSLSRASMTSTPPPPVPPPP 357
D 263 -----SOSSPSQHSPI-NSYTPPPPP-----LNTSTPSSSSFOGR 298
QY 358 PPATALQAPVPPPPAPLQIAPGVHPPPIAPLVQPSPPVARAACPVCETVPVHPLPQ 417
D 299 PPSGCFN---TPPPMSNN-----NNMPP--PPMQONGCAANNRLSVHNSAP----- 341
QY 418 GEVOGLPPPPPPPLPPGIRPSPVTVTALAHPPSGLHPTPSTAGPHVPLMPPSPSQ 477
D 342 --IVGAPAPPPPP-----PPS-----APAP-----PPPPMAK 367
QY 478 V-IPASEPKRHPSTPLVYSARSVLLEAIRKGIOLRVKEEQREQEAHERIENDVATILS 536
D 368 AGGASDIK-----PKASGARDLLSSIMOGMALKPAERKVAEPKKEALNVADILA 421
QY 537 RRIA-VEYSDSEDDSEFDEVDW 557
D 422 RRIAWAGDSSES-ESDSDM 442

RESULT 10
Q9TOK5 PRELIMINARY; PRT; 760 AA.
AC Q9TOK5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN T9E8.80 OR AT4G13340.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 348-747 FROM N.A.
RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```

Query Match	12.0%;	Score 358.5;	DB 10;	Length 409;			
Best Local Similarity	37.9%;	Pred. No. 6e-17;					
Matches	85;	Conservative.	9;	Mismatches 99; Indels 31; Gaps			
QY	275	PHEPPPPPMHAGADAKPIPTCISSATGLIENRQSPATGRTPVEVSPTPPPPPPLPSA 334					
Db	63	PEHPPPPPPPPPPPPPQPLPP-----SPSPPPPPPPPPPPPPPPPPPPPPPPSP 111					
QY	335	LSTSSLSRASMTSTPP 394					
Db	112	PP 164					
QY	395	QSPSPVARAACPVCETVPVHPLQGEVQGLPPPPPPPPPPPPPPPPPPPPPPPPPP 454					
Db	165	PPSPPPSP 217					
QY	455	LHPT---PSTAGCPHVLMP---PSPSQVIPASEKRPSTLP 492					
Db	218	RPSPSPSPSPPPSP 261					
RESULT	12						
ID	P70433	PRELIMINARY:	PRT:	802 AA.			
AC	P70433;	P70430;	P70431;	P70432;			
DT	01-FEB-1997	(TReMBLrel. 02, Created)					
DT	01-FEB-1997	(TReMBLrel. 02, Last sequence update)					
DE	01-WAR-2001	(TReMBLrel. 16, Last annotation update)					
DE	ENABLED	HOMOLOG.					
GN	ENAH	OR MENA.					
OS	Mus musculus	(Mouse).					
OS	Eukaryota;	Metazoa;	Chordata;	Craniala;	Vertebrata;	Euteleostomi;	
OX	Mammalia;	Eutheria;	Rodentia;	Sciurognathi;	Muridae;	Murinae;	Mus.
NCBI_TaxId	10090;						
LN	[1]						
RP	SEQUENCE FROM N.A.,	AND CHARACTERIZATION.					
RC	TISSUE=BRAIN;						
RA	MEDLINE=97015079;	PubMed=8861907;					
RT	Gerlter F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.;						
RT	"Mena, a relative of VASP and Drosophila Enabled, is implicated in the						
RT	control of microfilament dynamics.";						
RL	Cell 87:227-239(1996).						
CC	-!	FUNCTION: MAY BE INVOLVED IN MICROFILAMENT ASSEMBLY AND CELL					
CC	MOTILITY. INDUCES THE FORMATION OF F-ACTIN RICH OUTGROWTHS IN						
CC	FIBROBLASTS. BINDS PROFILIN.						
CC	-!	SUBCELLULAR LOCATION: LOCALIZED TO FOCAL ADHESIONS AND, TO A					
CC	LESSER EXTENT, LEADING EDGES AND STRESS FIBERS.						
CC	-!	ALTERNATIVE PRODUCTS: FOUR FORMS; MENA, MENA+, MENA++ AND MENA+++					
CC	(SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.						
CC	EMBL; U72520;	AAC52863.1;					
DR	EMBL; U72521;	AAC52864.1;					
DR	EMBL; U72522;	AAC52865.1;					
DR	EMBL; U72523;	AAC52866.1;					
DR	MGI; MGI:108360;	Enah.					
DR	InterPro; IPR000156;						
DR	InterPro; IPR000697;						
DR	InterPro; IPR001960;						
DR	Pfam; PF00568;	WH1; 1.					
DR	SMART; SM00160;	RanBD; 1.					
KW	Alternative splicing;	Phosphorylation.					
FT	DOMAIN	442	464				
FT	DOMAIN	542	552				
FT	DOMAIN	562	574				
FT	DOMAIN	578	589				
FT	DOMAIN	593	605				
FT	DOMAIN	686	689				
FT	MOD_RES	255	255				
FT	FT						
FT	FT						
FT	MOD_RES	637	637				
FT	FT						
FT	VARSPIC	117	131				
FT	VARSPIC	117	135				
FT	VARSPIC	132	135				
FT	FT						

FT VARSPLIC 259 500 MISSING (IN ISOFORM MENA).
SQ SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;

Query Match 12.0%; Score 356.5; DB 11; Length 802;
Best Local Similarity 22.5%; Pred. No. 1.7e-16;
Matches 151; Conservative 83; Mismatches 196; Indels 241; Gaps 28;

QY 3 LVKRN1-DPRHLCHTALPRGKNELECVTNLSLNIIRQLSSL-----SKYAEIDFG----- 53
DB 48 VVGRKIQRHVQVINCIPKGLKYN-QATQTHQRDARQVGLNFGSKEDANVFASAMMH 106
QY 54 --ELFNEAHSFVRNSQERDRLSV-----SVTQLDKKEBELSLQDITM 97
DB 107 ALEVLNSQEAQSKVTATQDSTNLRCIFCGPTLPNLSQPAQVQNGSQEELERQRL 166
QY 98 KFAFRSSIQOQDLDRKTLPIQETVDVCEQPLMLTPYRDDGKGLKFTYNNPYF 157
DB 167 OEQQRKELEREM-ERERLE-----RERLE----- 191
QY 158 FDLWKEKMLQD-TEDKRKREKROKKNLDRPHEP--EKVPRAPHDRRR-----E 203
DB 192 ----RERLERLEQERQERHERLERERLERERLERERLERERLERERLEQERQVE 247
QY 204 WOKLAQGPPELA-EDDANLLH-----KHIEVANGPAS----- 233
DB 248 WERERRRMAAPSSDSSLAPLPEYSSCQPPSPAPPSYAKVISAPVSDATPDYAVVTAL 307
QY 234 -----HFETRPTQTYVDHMDGYSLSALPFSQMS 262
DB 308 PPTSTPTPTPLRHAATRFATSLGSAFHPVLPHYATVPRPLNKNRSPSPVNT-PSQOPPA 366
QY 263 LLTRAERVLVRPHEPPPP-----PPMHGAGDAKPICTCISSATGLIENRPOSPATGR-- 315
DB 367 AKSCAMPNSFPLPPSPIMISSPGKATGRPVLPVCVSSP---VPQMPPSPPTAPNGS 423
QY 316 -----TPVFVSPT-----PPPPPPPLPSALSTSSLRASMTSTPPPPVPPPPPPATALQ- 364
DB 424 LDSVTYVSPPTSGAAPPPPPPPPP-----PPPPPPPLPPLPLPLSLH 470
QY 365 --APAVPPPPAPLQATPGVLHPAPPPPIAPPLVQSPVVARAAPVCETVPVH----- 413
DB 471 CGSQASPPPGTPL-----ASTPSKPSVLPSP--SAGAPASAEPTLPNPELGDSAS 519
QY 414 -----PLPQCEVOGLPPPPPPPL-----PPGRPSPTPTVTALA 449
DB 520 EPGLOAASQPAESTPQGLVGPAPPPPPPLPSPGAVASALPPPPGPPPPPLPSTGPP 579
QY 450 HPPSGLHPTPTAGPHVPLMPPPPSPQIVPASEPKRHPSTLPLVISDARS-----VILE 503
DB 580 PPPPPPPPLPNOAPP-----PPPPAPPLPAS-----GIFSGSTSEDNRPLTGLA 625
QY 504 AIRKGIQLRKV 514
DB 626 AAIAGAKLRKV 636

RESULT 13

Q9FPQ6 PRELIMINARY; PRT; 555 AA.

AC Q9FPQ6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VEGETATIVE CELL WALL PROTEIN GPI.
GN GPI.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI_TaxID=3055;

RN [1]

RP SEQUENCE FROM N.A.

RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,

RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RL plant hydroxyproline-rich glycoproteins.";
DR Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
SQ EMBL; AF309494; AAG45420.1;
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 11.9%; Score 354.5; DB 10; Length 555;
Best Local Similarity 38.3%; Pred. No. 1.6e-16;
Matches 92; Conservative 14; Mismatches 95; Indels 39; Gaps 8;

QY 275 PHEPPPPPMHAGDAKPICTCISSATGLIENRPOSPA--TGRTPVFVSPTPPPPPPPLP 332
DB 118 PPSPPSPAPPSPPAPPSPPSPAPPL-----PPSPAPPSPPSPPPVPPSPAP 173
QY 333 SALSTSSLRASMTSTPPPPVPPPPPPPPATALQAPVPPPPAPLQIAPGV----- 381
DB 174 P-----SPTTSPSPSPVPPSPAPSPAPVPPSPAPSPAPVPPSPAPSPSPS 222
QY 382 -LHPAPPPIAPPLVQSPVVARAAPVCETVPVHPLPQCEVOGLPPPPPPPLPPIGIRP- 439
DB 223 PAPSPSPAPPP--SPSPAPPSPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 280
QY 440 --SSPVTVTALAHPPSGLHPTPTSTAPGHVPL-MPPS-----PPSQVIPASEPKRHPSTLP 492
DB 281 PANTPMPPSPSPSPAPPTPTTPTTSPSPSPSPVPPSPAPVPPSPAPSPAPSPAPSPAP 340

RESULT 14

O00401 PRELIMINARY; PRT; 505 AA.

AC O00401;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE N-WASP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=97464048; PubMed=93272739;
RA Fukuoka M., Miki H., Takenawa T.;
RT "Identification of N-WASP homologs in human and rat brain.";
RL Gene 196.43-48(1997).
DR EMBL; D88460; BAA20128.1;
DR InterPro; IPR000095;
DR InterPro; IPR000697;
DR InterPro; IPR001960;
DR InterPro; IPR003124;
DR Pfam; PF00568; WH1; 1.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00246; WH2; 1.

SQ SEQUENCE 505 AA; 54824 MW; BC5670A11AB63539 CRC64;

Query Match 11.8%; Score 350.5; DB 4; Length 505;
Best Local Similarity 25.5%; Pred. No. 2.6e-16;
Matches 123; Conservative 53; Mismatches 146; Indels 161; Gaps 21;

QY 151 YTNYSYFEDLWKEKMLQ-----DTEDKRKKRK-----OKOKNLRDPEKVPKRAP- 197
DB 100 YNSPRGYFHTFAGTCCQVALNFANEEAKKFKAVTDLGLRRQRKRRRRPPNGNLP 159
QY 198 -----HRRRREWQKLAQGPPELAEDD-----ANLLKHIEVAN 229
DB 160 ATVDIKNPEITNRFYGVQVNNISHTREKKKGK-AKKRLTKGDIGTPSNFQH----- 211
QY 230 GPASHFETRQTYVD-----HMDGYSLSALPFSQMSSELLTRAERVL----- 272

Qy	415	LPQGEVGLPPPPPPPLPPPGIRPPSSSVTVTALAHPPSGLHPTSTAPCGPHVPLMPSP	474
Db	376	-PPPPASSPPPPPPPPSP-PPPPPPATAANPPSPA-PSRSRAGGEPPLGTRPPPP	430
Qy	475	PSQ	477
Db	431	PPE	433

Search completed: June 20, 2001, 12:10:37
Job time: 206 sec

```

RESULT 15
QI1645 ID Q41645 PRELIMINARY; PRT; 464 AA.
AC Q41645;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE EXTENSIN (FRAGMENT).
GN 15G.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=HK 10 (FEMALE) 69-1B (MALE);
RX MEDLINE=92289669; PubMed=1600938;
RA Ertl H., Hallmann A., Wenzl S., Sumper M.;
RT "A novel extensin that may organize extracellular matrix biogenesis in
Volvox carteri.";
RL EMBO J. 11:2053-2062(1992).
DR EMBL; X65165; CAA46283.1; -.
DR Mendel; 17416; Volca; 2658; 17416.
DR InterPro; IPR000480; -.
DR InterPro; IPR002965; -.
DR PRINTS; PR00211; GLUTELIN.
DR PRINTS; PR01217; PRICHEXTENSIN.
SQ NON_TER 1 1
FT SEQUENCE 464 AA; 47838 MW; 4228E3FDD53F70FO CRC64;

```

	Query Match	11.74;	Score 348.5;	DB 10;	Length 464;
	Best Local Similarity	37.0%;	Pred. No.	3.3e-16;	
	Matches	90;	Conservative	19;	Mismatches 89; Indels 45; Gaps 10;
Qy	263 LLTRAEEVLVRHEPP-----PPPDMHCAGDAKPI-----PTCIGSATGLIENRRQS	310 : : :			
Dd	208 LVTAADVVTSPGPPPRVSTSGPVPARSSSPPATRSPPPRRTTSPGVLTASPLL	267 			
Qy	311 PATG-RTEVFVSPPRP-----PPPLPSALSTSLLRASMTTPppvpvpp-----pppppa	360 : : : :			
Dd	268 PKTSPPPPRVPSPGPPPVASPPPPPPRVSPSPPPQPQVSPPPPPPPSRSPSPPPPRS	327 : : : :			
Qy	361 TALQAIVAPDPAPPPLQIAGVLHPAPP-----IAPLVGPSPVARAAPVCETVPVHP	414 : : : :			
Dd	328 SPSPSPSPSPSPSPSPSPPRS---SPSPGRSPSPSPSPSPSPSPSPSPSPSPSPSPSP	375 : : : :			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:07:11 ; Search time 37.19 Seconds
(without alignments)
1988.667 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 2981
Sequence: 1 MPLVKRNIDPRHLCHTALPR.....AVEYSDSEDDSEFDEVOWLE 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2942	98.7	559	11 Q9ERQ9	Q9erq9 mus musculus
2	1259.5	42.3	498	4 Q9Y6W5	Q9y6w5 homo sapien
3	1199	40.2	497	4 Q60794	Q60794 homo sapien
4	1198	40.2	496	4 Q9UDY7	Q9udy7 homo sapien
5	1196.5	40.1	502	4 Q9UFI6	Q9ufi6 homo sapien
6	1027.5	34.5	455	4 Q94974	Q94974 homo sapien
7	807	27.1	613	5 Q9VKM2	Q9vkm2 drosophila
8	598	20.1	507	5 Q9XVK6	Q9xvk6 caenorhabdi
9	449	15.1	443	5 Q9XVA8	Q9xva8 dictyosteli
10	363	12.2	760	10 Q9T0K5	Q9t0k5 arabidopsis
11	358.5	12.0	409	10 Q9SBM1	Q9sbm1 volvox cart
12	356.5	12.0	802	11 P70433	P70433 mus musculus
13	354.5	11.9	555	10 Q9PPQ6	Q9fpq6 chlamydomon
14	350.5	11.8	505	4 Q90401	Q90401 homo sapien
15	348.5	11.7	464	10 Q41645	Q41645 volvox cart
16	347	11.6	711	10 Q9SPM1	Q9spm1 lycopersico
17	346.5	11.6	727	10 Q9XIL9	Q9xil9 arabidopsis
18	342.5	11.5	501	11 Q08816	Q08816 rattus norv
19	340.5	11.4	956	10 Q9LJ64	Q9lj64 arabidopsis

ALIGNMENTS

RESULT 1	11.4	1289	10	Q9FLO7	Q9fliq7 arabidopsis
Q9ERQ9	11.2	1307	10	Q9LVN1	Q9lvn1 arabidopsis
ID Q9ERQ9	11.2	505	6	Q95107	Q95107 bos taurus
AC Q9ERQ9	11.0	1151	5	Q9VAY4	Q9vay4 drosophila
DT 01-MAR-2001 (TREMBLrel. 16, Created)	11.0	1253	13	Q9DEH3	Q9deh3 gallus gall
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	11.0	1315	10	Q9SPM0	Q9spm0 zea mays (m
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)	11.0	574	3	Q36027	Q36027 schizosacch
DE WAVE-1	10.9	1188	10	Q41805	Q41805 zea mays (m
OS Mus musculus (Mouse)	10.8	967	5	Q9NGX2	Q9ngx2 entamoeba h
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	10.8	1096	5	Q9NGX1	Q9ngx1 entamoeba h
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	10.7	1006	10	Q9LMQ1	Q9vnc2 drosophila
OX NCBI_TaxID=10090;	10.7	446	5	Q9WZC2	Q9wzh2 mycobacteri
RN [1]	10.7	763	2	Q9XDH2	Q9xhd2 arabidopsis
RP SEQUENCE FROM N.A.	10.6	699	10	Q81765	Q81765 arabidopsis
RX MEDLINE=20428428; PubMed=10970852;	10.6	631	10	Q9LT74	Q9lt74 arabidopsis
RA Westphal R.S., Soderling S.H., Alto N.M., Langeberg L.K., Scott J.D.;	10.6	744	10	Q65375	Q65375 arabidopsis
"Scar/WAVE-1, a wiskott-aldrich syndrome protein, assembles an actin-	10.3	451	11	Q62775	Q62775 rattus norv
RT associated multi-kinase scaffold.";	10.3	485	11	Q9Z0G8	Q9z0g8 rattus norv
RL EMBO J. 19:4589-4600(2000).	10.2	708	10	Q9SX31	Q9sx31 arabidopsis
DR EMBL AF290877; AAC02214.1	10.2	786	10	Q48809	Q48809 arabidopsis
SQ SEQUENCE 559 AA; 61508 MW; 8746910987D80D16 CRC64;	10.2	1012	4	Q43393	Q43393 homo sapien
	10.2	1006	11	Q62901	Q62901 rattus norv
	10.1	1012	4	Q75359	Q75359 homo sapien
	10.1	3247	14	Q65553	Q65553 bovine hecp
	10.1	520	11	Q61078	Q61078 mus musculu
	10.0	1566	4	Q9P2R6	Q9p2r6 homo sapien

Query Match	98.7%;	Score 2942;	DB 11;	Length 559;
Best Local Similarity	98.4%;	Pred. No. 5e-193;		
Matches 550;	Conservative	3;	Mismatches	6;
			Indels	0;
			Gaps	0;
Qy 1	MPLVKRNIDPRHLCHTALPRGKNELECVTNISLANIIRQLSSLSKYAEDIFGELFNEAH	60		
Db 1	MPLVKRNIDPRHLCHTALPRGKNELECVTNISLANIIRQLSSLSKYAEDIFGELFNEAH	60		
Qy 61	SFSPRVNSLOERDRLSVSVTOLDPKKEELSLODITWRKAFRSSTIODQOLFDRKTLPIP	120		
Db 61	SFSPRVNSLOERDRLSVSVTOLDPKKEELSLODITWRKAFRSSTIODQOLFDRKTLPIP	120		
Qy 121	LOETDYVCEOPPPPLNLTPTPRDDGKEGLKFTYNPSPFFDLWKEKMLQDTEDEKREKRRQK	180		
Db 121	LOETDYVCEOPPPPLNLTPTPRDDGKEGLKFTYNPSPFFDLWKEKMLQDTEDEKREKRRQK	180		
Qy 181	QKNLDRPHEPEKVPRAHPDRRRREWKQAQPELAEDDANLLHKHIEVANGPASHFETRPQ	240		
Db 181	QKNLDRPHEPEKVPRAHPDRRRREWKQAQPELAEDDANLLHKHIEVANGPASHFETRPQ	240		

Qy	241	TYVDHMDGYSLSALPFSQMSSELLTRAERVLVRPHEPPPPPMMHGAGDAKPIPTCISSA	300
Db	241	TYVDHMDGYSLSALPFSQMSSELLTRAERVLVRPHEPPPPPMMHGAGDAKTPTCISSA	300
Qy	301	TGLIENRPOSATGRTPVFSVSTPTTTPPPPLPSALTSSLSRASMTSTPTPPVPVPPPPPPA	360
Db	301	TGLIENRPOSAPAGRTPVFSVSTPTTTPPPPLPSALTSSLSRASMTSTPTPPVPVPPPPPPA	360
Qy	361	TALQAAPAVPPPPAPLQIAPGVHLHPAPPTIAPPLVQPSPPVARAAPVCETVPVHPLPQGEV	420
Db	361	TALQAAPAVPPPPAPLQIAPGVHLHPAPPTIAPPLVQPSPPVARAAPVCETVPVHPLPQGEV	420
Qy	421	OQLPPTPPPPPLPPPCGIRSSPVVTYALAHPPSGLHTPTSTAPGHVPLMPSPSPSQVIP	480
Db	421	OQLPPTPPPPPLPPPGCGIRSSPVAAALAHPPSGHLHPADSTAPGPHAPLMPPSPSPSVLP	480
Qy	481	ASEPKRHPSTLPVISDARSVLLEAIRKGIQLRKVVEEQREQAEAKHERIENDVATILSRRIA	540
Db	481	ASEPKRHPSTLPVISDARSVLLEAIRKGIQLRKVVEEQREQAEAKHERIENDVATILSRRIA	540
Qy	541	VEYSDSEDDSEFDEVDWLE 559	
Db	541	VEYSDSEDDSEFDEVDWLE 559	

RESULT 2

QY6W5	ID	QY6W5	PRELIMINARY;	PRT;	498 AA.
AC	QY6W5;				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DE	01-NAR-2001	(TREMBLrel. 16, Last annotation update)			
DE	WASP-FAMILY PROTEIN	(DJ144C9.2).			
GN	WAVE2 OR DJ144C9.2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBII_Taxid=9606;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=99310608; PubMed=10381382;				
RA	Suetsugu S., Miki H., Takenawa T.				
RT	"Identification of two human WAVE/SCAR homologues as general actin				
RL	regulatory molecules which associate with the Arp2/3 complex.";				
RL	Biochem. Biophys. Res. Commun. 260:296-302(1999).				
RN	[2]				
RN	SEQUENCE FROM N.A.				
RN	Hall R.;				
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL; AB026542; BAA81795.1; -				
DR	EMBL; AL096774; CAC18518.1; -				
DR	InterPro; IPR002965; -				
DR	InterPro; IPR002966; -				
DR	InterPro; IPR003124; -				
DR	Pfam; PF02205; WH2; 1.				
DR	PRINTS; PR01217; PRICEXTENSIN.				
DR	SMART; SM00246; WH2; 1.				
SQ	SEQUENCE 498 AA: 54283 MW; C737C9B63016D94 CRC64;				

[illegible]

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=93041923; PubMed=1420303;
RA Sazuka T., Tomooka Y., Kathju S., Ikawa Y., Noda M., Kumar S.;
RT "Identification of a developmentally regulated gene in the mouse
RL Biochim. Biophys. Acta 1132:240-248(1992).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT BIOLOGICAL ROLE IN THE
CC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM.
CC -1- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
CC LONG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10727; BAA01570.1;
DR MGD; MGI:97291; Ndppl.
KW Developmental protein.
FT DOMAIN 30 52 POLY-PRO.
FT DOMAIN 163 195 LEU/PRO-RICH.
FT DOMAIN 273 276 POLY-ARG.
FT SEQUENCE 389 AA; 41249 MW; EE36C1CBEB8156033 CRC64;

Query Match 2.1%; Score 12; DB 1; Length 389;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
DB 37 PPPPPPPPLPPP 48
|||||
-----
RESULT 4
Y009_HUMAN STANDARD; PRT; 314 AA.
AC Q15390;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0009.
GN KIAA0009.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow.
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1."
RL DNA Res. 1:27-35(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: D13634; BAA02798.1;
KW Hypothetical protein.
FT DOMAIN 165 178 POLY-PRO.
SQ SEQUENCE 314 AA; 34756 MW; E79E4F4E310619E3 CRC64;

Query Match 2.0%; Score 11; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 0.0069;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 PPPPPPPPLPPP 435
DB 168 PPPPPPPPLPPP 178
|||||
-----
RESULT 5
CAP2_HUMAN STANDARD; PRT; 477 AA.
AC P40123;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENYLYL CYCLASE-ASSOCIATED PROTEIN 2 (CAP 2).
GN CAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95051124; PubMed=7962207;
RA Yu G., Swiston J., Young D.;
RT "Comparison of human CAP and CAP2, homologs of the yeast adenylyl
RT cyclase-associated proteins."
RL J. Cell Sci. 107:1671-1678(1994).
CC -1- FUNCTION: MAY HAVE A REGULATORY BIFUNCTIONAL ROLE.
CC -1- SUBCELLULAR LOCATION: CELL MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CAP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02390; AAA20587.1;
DR InterPro; IPR001837;
DR Pfam; PF01213; CAP; 1.
DR PROSITE; PS01088; CAP_1; 1.
DR PROSITE; PS01089; CAP_2; 1.
KW Membrane; Multigene family.
FT DOMAIN 232 245 POLY-PRO.
FT SEQUENCE 477 AA; 52824 MW; 2FD3950C094F5AF7 CRC64;

Query Match 2.0%; Score 11; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PPPPPPPPLPPP 436
DB 232 PPPPPPPPLPPP 242
|||||
-----
RESULT 6
Y029_HUMAN STANDARD; PRT; 971 AA.
AC Q15032;
DT 15-JUL-1998 (Rel. 36, Created)
DT 13-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
CC -----
```

DE HYPOTHETICAL PROTEIN KIAA0029.
GN KIAA0029.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.";
RL DNA Res. 1:27-35(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21852; BAA04878.1; -;
DR InterPro; IPR001374; -;
DR Pfam; PF01424; R3H; 1.
KW Hypothetical protein.
FT DOMAIN 451 480
FT DOMAIN 461 480
FT DOMAIN 694 698
FT POLY-PRO.
FT POLY-PRO.
SQ SEQUENCE 971 AA; 107673 MW; D07684D368955108 CRC64;

Query Match 2.0%; Score 11; DB 1; Length 971;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPP 434
DB 470 PPPPPPPPLPP 480

RESULT 7
CLC1_RAT
ID CLC1_RAT STANDARD; PRT; 994 AA.
AC P35524;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHLORIDE CHANNEL PROTEIN, SKELETAL MUSCLE (CHLORIDE CHANNEL PROTEIN
DE 1) (CLC-1).
GN CLCNI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=92065954; PubMed=1659664;
RA Steinmeyer K., Ortlad C., Jentsch T.J.;
RT "Primary structure and functional expression of a developmentally
RT regulated skeletal muscle chloride channel.";
RL Nature 354:301-304(1991).
CC -1- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLES.
CC -1- SIMILARITY: TO OTHER CHLORIDE CHANNELS.

CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62894; CAA44683.1; -;
DR PIR; S19595; S19595.
DR InterPro; IPR000644; -;
DR InterPro; IPR001807; -;
DR InterPro; IPR002243; -;
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF00654; voltage_CLC; 1.
DR PRINTS; PR00762; CLCHANNEL.
DR PRINTS; PR01112; CLCHANNEL.
KW Ionic channel; Ion transport; Voltage-gated channel; Transmembrane;
KW CBS domain.
FT DOMAIN 1 117
FT TRANSMEM 118 137
FT TRANSMEM 159 182
FT TRANSMEM 206 227
FT TRANSMEM 235 254
FT TRANSMEM 266 290
FT TRANSMEM 305 323
FT TRANSMEM 348 368
FT TRANSMEM 391 414
FT TRANSMEM 456 475
FT TRANSMEM 478 496
FT TRANSMEM 524 545
FT TRANSMEM 553 572
FT DOMAIN 573 844
FT TRANSMEM 845 863
FT DOMAIN 864 994
FT DOMAIN 607 662
FT DOMAIN 825 877
FT CBS 1.
FT CBS 2.
SQ SEQUENCE 994 AA; 110073 MW; DCDDCD0D26E48FAE CRC64;

Query Match 2.0%; Score 11; DB 1; Length 994;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPP 434
DB 728 PPPPPPPPLPP 738

RESULT 8
NO20_SOYBN
ID NO20_SOYBN STANDARD; PRT; 185 AA.
AC P08960;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NODULIN 20 PRECURSOR (N-20).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87146477; PubMed=3822835;
RA Sandal N.N., Bojsen K., Marcker K.A.;
RT "A small family of nodule specific genes from soybean.";
RL Nucleic Acids Res. 15:1507-1519(1987).
RN [2]
RP SEQUENCE FROM N.A. (TRUNCATED FORM N-20T).
RC STRAIN=CV. DARE; TISSUE=Leaf;

RX MEDLINE-92049245; PubMed-1840639;
RA Richer H.E., Sandal N.N., Marcker K.A., Sengupta-Gopalan C.;
RT "Characterization and genomic organization of a highly expressed late
RL nodulin gene subfamily in soybeans.";
RL Mol. Gen. Genet. 229:445-452(1991).
CC -1- SUBCELLULAR LOCATION: PERIBACTEROID MEMBRANE OR PERIBACTEROID
CC -1- SPACE (POTENTIAL).
CC -1- INDUCTION: DURING NODULATION IN LEGUME ROOTS AFTER RHIZOBIUM
CC INFECTON.
CC -1- MISCELLANEOUS: THE AUTHORS SUGGEST THAT THE 2 CONSERVED DOMAINS
CC EACH INCLUDING TWO CYC-TAA-CYS ARRANGEMENTS ARE METAL-BINDING
CC DOMAINS, ANALOGOUS TO THE 'FINGER REGIONS' PRESENT IN
CC METAL-BINDING PROTEINS; THE BINDING OF METAL IONS SEEMS TO BE
CC IMPORTANT FOR THE FUNCTION OF THESE NODULINS.
CC -1- SIMILARITY: TO NODULIN 22, 23 AND 44 AND PARTIAL TO METAL-BINDING
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X05020; CAA28677.1; -;
DR EMBL; X60159; CAA42728.1; -;
DR PIR; B26669; B26669.
KW Nodulation; signal; Metal-binding.
FT SIGNAL 1 17
FT CHAIN 18 185
FT SIMILAR 38 94
FT TO NODULIN 22, 23, 44 AND TO THE 'FINGER
FT REGIONS' OF METAL-BINDING PROTEINS.
FT TO NODULIN 22, 23, 44 AND TO THE 'FINGER
FT REGIONS' OF METAL-BINDING PROTEINS.
FT MISSING (IN TRUNCATED FORM N-20T).
FT VARIANT 76 185
FT SEQUENCE 185 AA; 20022 MW; 94F21F90B68A6A2A2 CRC64;
SQ

Query Match 1.8%; Score 10; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 LPPPPPPPL 432
Db 175 LPPPPPPPL 184

RESULT 9

ID YN03_CAEEL STANDARD; PRT; 211 AA.
AC Q03607;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOHETICAL 23.6 KDA PROTEIN T23G5.3 IN CHROMOSOME III.
GN T23G5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latrelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; Z19158; CAA79571.1; -;
DR PIR; S28304; S28304.
DR WormPep; T23G5.3; CE00333.
KW Hypothetical protein.
KW SEQUENCE 211 AA; 23552 MW; 85D829BDEA17740 CRC64;
SQ

Query Match 1.8%; Score 10; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPL 433
Db 90 PPPPPPPPL 99

RESULT 10

ID SMN_MOUSE STANDARD; PRT; 288 AA.
AC P97801; O09092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SURVIVAL MOTOR NEURON PROTEIN.
GN SMN1 OR SMN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97224505; PubMed-9070939;
RA Viollet L., Bertrand S., Brunialti A.L.B., Lefebvre S., Buriel P.,
RA Clermont O., Cruaud C., Guenet J.-L., Munnich A., Melki J.;
RT "cDNA isolation, expression, and chromosomal localization of the
RL mouse survival motor neuron gene (Smn).";
RL Genomics 40:185-188(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97264340; PubMed-9110173;
RA Didonato C.J., Chen X.N., Noya D., Korenberg J.R., Nadeau J.H.,
RA Simard L.R.;
RT "Cloning, characterization, and copy number of the murine survival
RT motor neuron gene: homolog of the spinal muscular atrophy-determining
RL gene.";
RL Genome Res. 7:339-352(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BALB/C; TISSUE-Brain;
RX MEDLINE-97420785; PubMed-9275227;
RA Schrank B., Goetz R., Gunnarsen J.M., Ure J.M., Toyka K.V.,
RA Smith A.G., Sendtner M.;
RT "Inactivation of the survival motor neuron gene, a candidate gene for
RT human spinal muscular atrophy, leads to massive cell death in early
RT mouse embryos.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9920-9925(1997).
CC -1- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICEOSOMAL
CC SNRNP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA

CC SPlicing IN THE NUCLEUS (BY SIMILARITY).
CC -!- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR
CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND
CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN
CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES
CC (SNRNPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPLICEOSOMAL
CC SNRP CORE SM PROTEINS, INCLUDING B/B', SM D AND SM E PROTEINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOSOLIC AND NUCLEAR. LOCALIZED IN
CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMS, WHICH
CC ARE HIGHLY ENRICHED IN SPLICEOSOMAL SNRNPs (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SMN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U63294; AAC53057.1; -
CC EMBL: U77714; AAC53144.1; -
CC EMBL: Y12835; CAA73356.1; -
CC MGD; MGI:109257; Smn.
KW mRNA processing; RNA-binding; Nuclear protein.
FT DOMAIN 190 196 POLY-PRO.
FT DOMAIN 212 222 POLY-PRO.
FT DOMAIN 239 243 POLY-PRO.
SQ SEQUENCE 286 AA; 31254 MW; 757B3074649F7458 CRC64;

Query Match 1.8%; Score 10; DB 1; Length 288;
Best Local Similarity 100.08; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PPPPPPLPPP 435
Db 212 PPPPPPLPPP 221
|||||

RESULT 11
ID P53_CHICK STANDARD; PRT; 367 AA.
AC P10360;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (TUMOR SUPPRESSOR P53).
GN TP53.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPAFAS;
RX MEDLINE=89083584; PubMed=3060861;
RA Soussi T.;
RT "Nucleotide sequence of a cDNA encoding the chicken p53 nuclear
RT oncoprotein.";
RL Nucleic Acids Res. 16:11383-11383(1988).
CC -!- FUNCTION: ACTS AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELL DIVISION BY
CC CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE
CC ACTIVATED GENES IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).

CC -!- SUBUNIT: BINDS DNA AS AN HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X13057; CAA31456.1; -
CC PIR: S02193; S02193.
CC HSP; P04637; lTsr.
CC InterPro: IPR002117; -
CC Pfam: PF00870; P53; 1.
CC PRINTS: PR00386; P53SUPPRESSR.
CC PROSITE: PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 30 TRANSCRIPTION ACTIVATION (ACIDIC).
FT DNA_BIND 87 278 BY SIMILARITY.
FT DOMAIN 308 339 OLIGOMERIZATION.
FT DOMAIN 347 364 BASIC (REPRESSION OF DNA-BINDING).
FT DOMAIN 292 306 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 366 366 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 367 AA; 40169 MW; FC37D0FCDF9195B6 CRC64;

Query Match 1.8%; Score 10; DB 1; Length 367;
Best Local Similarity 100.08; Pred. No. 0.075;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 PPPPPPLPPP 433
Db 50 PPPPPPLPPP 59
|||||

RESULT 12
CAP2_RAT STANDARD; PRT; 477 AA.
ID CAP2_RAT
AC P52481;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ADENYL CYCLASE-ASSOCIATED PROTEIN 2 (CAP).
GN CAP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96096534; PubMed=8522189;
RA Swiston J., Hubbard A., Yu G., Young D.;
RT "Differential expression of CAP and CAP2 in adult rat tissues.";
RL Gene 165:273-277(1995).
CC -!- FUNCTION: MAY HAVE A REGULATORY BIFUNCTIONAL ROLE.
CC -!- SUBCELLULAR LOCATION: CELL MEMBRANE (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: FOUND AT RELATIVELY HIGH LEVELS IN TESTES, AT
CC MODERATE LEVELS IN BRAIN, HEART AND SKELETAL MUSCLE, AT LOWER
CC LEVELS IN LUNG, SKIN, KIDNEY AND SMALL INTESTINE, AND IS
CC UNDETECTABLE IN LIVER OR SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE CAP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U31935; AAA92298.1; -
DR InterPro; IPR001837; -
DR Pfam; PF01213; CAP; 1.
DR PROSITE; PS01088; CAP_1; 1.
DR PROSITE; PS01089; CAP_2; 1.
KW Membrane; Multigene family.
FT DOMAIN 233 247 POLY-PRO.
SQ SEQUENCE 477 AA; 52912 MW; CFCDS58F256655B CRC64;

Query Match 1.8%; Score 10; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 0.096; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 GLPPPPPPPP 431
DB 231 GLPPPPPPPP 240
|||||

RESULT 13
CAP_SCHPO STANDARD; PRT; 551 AA.
ID CAP_SCHPO STANDARD; PRT; 551 AA.
AC P36621;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-JUN-2000 (Rel. 39, Last annotation update)
DE ADENYL CYCLASE-ASSOCIATED PROTEIN (CAP).
GN CAP OR SPCC306.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MK141;
RX MEDLINE=92199347; PubMed=1550959;
RA Kawamukai M., Gerst J., Field J., Riggs M., Rodgers L., Wigler M.,
Young D.;
RT "Genetic and biochemical analysis of the adenylyl cyclase-associated
protein, cap, in Schizosaccharomycetes pombe.";
RL Mol. Biol. Cell 3:167-180(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE N-TERMINAL DOMAIN BINDS TO ADENYL CYCLASE, THEREBY
CC ENABLING ADENYL CYCLASE TO BE ACTIVATED BY UPSTREAM REGULATORY
CC SIGNALS, SUCH AS RAS. THE C-TERMINAL DOMAIN IS REQUIRED FOR NORMAL
CC CELLULAR MORPHOLOGY AND GROWTH CONTROL.
CC -!- SUBCELLULAR LOCATION: CELL MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CAP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L16577; AAA35292.1; -
DR EMBL; AL049728; CAB41657.1; -
DR PIR; A60047; A60047.
DR InterPro; IPR001837; -
DR Pfam; PF01213; CAP; 1.
DR PROSITE; PS01088; CAP_1; 1.
DR PROSITE; PS01089; CAP_2; 1.
KW Membrane.
FT DOMAIN 288 315 ALA/PRO/SER-RICH.
FT DOMAIN 306 314 POLY-PRO.

SQ SEQUENCE 551 AA; 60243 MW; 2D7E82A953B1339E CRC64;

Query Match 1.8%; Score 10; DB 1; Length 551;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 GLPPPPPPPP 431
DB 304 GLPPPPPPPP 313
|||||

RESULT 14
CN4D_HUMAN STANDARD; PRT; 809 AA.
ID CN4D_HUMAN STANDARD; PRT; 809 AA.
AC Q08499; Q13549; Q13550; Q13551; O43433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CAMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE 4D (EC 3.1.4.17)
DE (PDE3) (PDE43).
GN PDE4D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (FORMS 1/HPDE4D3 AND 2/HPDE4D4).
RX MEDLINE=94019330; PubMed=8413254;
RA Bolger G., Michael T., Martins T., John T., Steiner B.,
Rodgers L., Riggs M., Wigler M., Ferguson K.;
RT "A family of human phosphodiesterases homologous to the dunce
learning and memory gene product of Drosophila melanogaster are
potential targets for antidepressant drugs.";
RL Mol. Cell. Biol. 13:6558-6571(1993).
RN [2]
RP SEQUENCE FROM N.A. (ALL FORMS), AND REVISIONS TO FORM 2/HPDE4D4.
RX MEDLINE=98041897; PubMed=9371713;
RA Bolger G.B., Erdogan S., Jones R.E., Loughney K., Scotland G.,
Hoffmann R., Wilkinson I., Farrell C., Houslay M.D.;
RT "Characterization of five different proteins produced by
alternatively spliced mRNAs from the human cAMP-specific
phosphodiesterase PDE4D gene.";
RL Biochem. J. 328:539-548(1997).
RN [3]
RP SEQUENCE FROM N.A. (FORM 3).
RC TISSUE=Heart;
RX MEDLINE=94171048; PubMed=8125310;
RA Baeker P.A., Oberholte R., Bach C., Yee C., Shelton E.R.;
RT "Isolation of a cDNA encoding a human rolipram-sensitive cyclic AMP
phosphodiesterase (PDE IVD).";
RL Gene 138:253-256(1994).
RN [4]
RP SEQUENCE FROM N.A. (FORMS 4/HPDE4D1; 5/HPDE4D2 AND 2/HPDE4D3).
RX MEDLINE=96390839; PubMed=8797812;
RA Nemoz G., Zhang R.B., Sette C., Conti M.;
RT "Identification of cyclic AMP-phosphodiesterase variants from the
PDE4D gene expressed in human peripheral mononuclear cells.";
RL FEBS Lett. 384:197-102(1996).
CC -!- CATALYTIC ACTIVITY: ADENOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O -
CC ADENOSINE 5'-PHOSPHATE.
CC -!- ENZYME REGULATION: INHIBITED BY ROLIPRAM.
CC -!- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.
CC -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF THE PROTEIN ARE PRODUCED
CC BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: WIDESPREAD; MOST ABUNDANT IN SKELETAL MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
DR EMBL; L20970; AAA03592.1; -
DR EMBL; L20969; AAC00042.1; -
DR EMBL; U02882; AAC13745.1; -
DR EMBL; U50157; AAA97890.1; -
DR EMBL; U50158; AAA97891.1; -
DR EMBL; U50159; AAA97892.1; -
DR EMBL; AF012074; AAC00070.1; -
DR EMBL; AF012073; AAC00069.1; -
DR MIM; 600129; -
DR InterPro; IPR002073; -
DR Pfam; PF00233; PDEase; 1.
DR PRINTS; PR00387; PDJESTERASE1.
DR PROSITE; PS00126; PDEASE_1; 1.
KW Hydrolase; cAMP; Alternative
FT VARSPLIC 1 152
FT MEAGSSAPARAGSGSDSAGGATLKAPKHLWRHEQHQQY
PLRQFRLPHHLLPPPPSPQPCPLQPPPPPLP
PPPPGGAARGYASSGATGVRHGRYSSTERYLYCRAMDR
TSYAVETGHRPGLAKSRMSWSPSQGLRR -> MAQQTSPD
(IN ISOFORM 6/HPDE4D3).
FT VARSPLIC 1 152
FT MEAGSSAPARAGSGSDSAGGATLKAPKHLWRHEQHQQY
PLRQFRLPHHLLPPPPSPQPCPLQPPPPPLP
PPPPGGAARGYASSGATGVRHGRYSSTERYLYCRAMDR
TSYAVETGHRPGLAKSRMSWSPSQGLRR -> MMHVNNFF
RHSWIC (IN ISOFORM 3).
FT VARSPLIC 1 205
FT MISSING (IN ISOFORM 3).
FT MEAGSSAPARAGSGSDSAGGATLKAPKHLWRHEQHQQY
PLRQFRLPHHLLPPPPSPQPCPLQPPPPPLP
PPPPGGAARGYASSGATGVRHGRYSSTERYLYCRAMDR
TSYAVETGHRPGLAKSRMSWSPSQGLRR -> MKEQPCAGTGHPM
AGYGRMAPFELASGPFVKRLRTSPFPCFLA (IN
ISOFORM 4/HPDE4D1).
FT VARSPLIC 1 302
FT MISSING (IN ISOFORM 5/HPDE4D2).
FT CONFLICT 644 644
FT R -> P (IN REF. 3).
FT CONFLICT 769 769
FT C -> R (IN REF. 4).
SQ SEQUENCE 809 AA; 91114 MW; 7A4773DD3A044F57 CRC64;
```

Query Match 1.8%; Score 10; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PPPPPPLPPP 435

Db 75 PPPPPPLPPP 84

RESULT 15

```
ID DIA3_HUMAN STANDARD; PRT; 853 AA.
AC Q9NSV4;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 3 (DIAPHANOUS-RELATED FORMIN 3) (DRF3)
DE (FRAGMENT).
GN DIAPH3 OR DIAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC Tissue=Testis;
RA Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE OF 468-853 FROM N.A. (ISOFORMS 1 AND 2).
RA Smith M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -----
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
DR EMBL; AL137718; CAB70890.1; ALT_INIT.
DR EMBL; AL354829; CAC17664.1; -
DR EMBL; AL354829; CAC17665.1; -
DR InterPro; IPR003104; -
DR Pfam; PF02181; FH2; 1.
KW Coiled coil; Repeat; Alternative splicing.
FT NON_TER 1
FT DOMAIN 304 374 FH1 (PRO-RICH).
FT DOMAIN 379 818 FH2.
FT DOMAIN 801 815 DAD.
FT DOMAIN 819 822 ARG/LYS-RICH (BASIC).
FT VARSPLIC 656 697
FT VSVETLEKNLRQMGRLQQLKELETPPPPPDLHDKFVTKI
FT F -> GLCLFKKHFALISAKRLKIIPFCIMYPLSHSVF
FT IPNISF (IN ISOFORM 2).
FT MISSING (IN ISOFORM 2).
SQ SEQUENCE 853 AA; 98610 MW; B7FA9C745AE18CD9 CRC64;
```

Query Match 1.8%; Score 10; DB 1; Length 853;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPLPPP 433

Db 334 PPPPPPLPPP 343

Search completed: June 20, 2001, 12:14:55
Job time: 194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:06:45 ; Search time 14.07 Seconds
(without alignments)
800.345 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 2981
Sequence: 1 MPLVKRNDPRHLCHTALPR.....AVEYSDSEDDSEFDEVDMLE 559

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320.5	10.8	1315	4	US-08-899-595-3
2	318.5	10.7	1255	2	US-09-080-897-4
3	318.5	10.7	1255	4	US-08-899-595-1
4	318.5	10.7	1255	4	US-09-323-735-4
5	313.5	10.5	1248	2	US-09-080-897-2
6	313.5	10.5	1248	4	US-09-323-735-2
7	261	8.8	214	1	US-08-217-327-4
8	259.5	8.7	306	1	US-08-217-327-6
9	246.5	8.3	297	2	US-08-580-545B-6
10	246.5	8.3	297	4	US-09-262-653A-6
11	246.5	8.3	1185	4	US-09-041-886-23
12	244	8.2	330	1	US-08-642-255-32
13	244	8.2	408	1	US-07-609-716-65
14	244	8.2	408	4	US-08-475-411A-65
15	244	8.2	408	4	US-08-478-029A-65
16	234	7.8	2414	1	US-08-227-536-2
17	234	7.8	2414	5	PCT-US95-04682-2
18	228	7.6	1333	4	US-09-356-952-2
19	227.5	7.6	1305	2	US-08-574-959A-9
20	227.5	7.6	1135	2	US-08-574-959A-7
21	224.5	7.5	2441	1	US-08-194-468-2
22	224.5	7.5	2441	3	US-08-961-739-2
23	217	7.3	334	6	5202236-3
24	212	7.1	331	6	5202236-37
25	209.5	7.0	1319	2	US-08-290-731C-2
26	209.5	7.0	1336	2	US-08-290-731C-6
27	208.5	7.0	1291	4	US-09-150-460B-10

28	208.5	7.0	1291	4	US-09-220-641-5	Sequence 5, Appli
29	208	7.0	311	4	US-09-179-558-66	Sequence 66, Appl
30	206	6.9	722	1	US-08-347-718B-1	Sequence 1, Appli
31	206	6.9	722	1	US-08-445-050-3	Sequence 3, Appli
32	206	6.9	722	1	US-08-445-050-7	Sequence 7, Appli
33	206	6.9	722	2	US-08-482-262-1	Sequence 1, Appli
34	206	6.9	722	2	US-08-204-691-3	Sequence 3, Appli
35	206	6.9	722	2	US-08-204-691-7	Sequence 7, Appli
36	206	6.9	722	6	5200183-3	Patent No. 5200183
37	206	6.9	742	1	US-08-347-718B-2	Sequence 2, Appli
38	206	6.9	742	2	US-08-482-262-2	Sequence 2, Appli
39	206	6.9	742	6	5200183-2	Patent No. 5200183
40	206	6.9	745	1	US-08-445-050-2	Sequence 2, Appli
41	206	6.9	745	2	US-08-204-691-2	Sequence 2, Appli
42	206	6.9	745	3	US-08-370-223-13	Sequence 13, Appli
43	203	6.8	960	4	US-09-219-849-5	Sequence 5, Appli
44	202.5	6.8	1162	2	US-08-728-323A-2	Sequence 2, Appli
45	201.5	6.8	357	1	US-07-609-716-66	Sequence 66, Appli

ALIGNMENTS

RESULT 1
US-08-899-595-3
; Sequence 3, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-595-3

Query Match 10.8%; Score 320.5; DB 4; Length 1315;

Best Local Similarity 25.4%; Pred. No. 1.3e-15;
Matches 148; Conservative 59; Mismatches 183; Indels 193; Gaps 34;

Qy 51 IFGELFNEAHFSFRVNSLQERV-----DRLSVSVTOLDPKEEELS--- 91
Db 360 LINALITPAEELDFRVHISRLMRLGLHQLVLOELREIENEDMRVOLNFVDEQGEEDSYDL 419
Qy 92 ---LQDITMRKAFFSSTIQDQOLFDRKTLPIPLQETVDCVQPPPLNLTLP---YRDGK 145
Db 420 KGRLLDIRMDDNEVQ-----ILLNTVSKAEPHFLSLOHLLVRNDYE 468
Qy 146 BGLKPY-----TNPYS---FFDLWKMKLQDTEKRR-EKRRKQKQKND 185
Db 469 ARPQYKLIBECISQVLHKNAGADPDFCRHLQIEGLIDQIMDKTKVKESEAKAAELE 528
Qy 186 RPHEPEKVPRAPHRRRWOKLAOGPELAEDDANLLKHIEVANGPASHFETRPQTYVDH 245
Db 529 KKLDSLTAR---HELQVEMKKHESFEOKLQDLQ-----GEKDALHSE-KQOIATEK 577
Qy 246 MDGYSLSALPFSOMSELLTRAEE-----RVLVRPHEP-----PPPPMHGAGD 289
Db 578 QDLEAEVSQLT-GEVAKLTKELEDAKKEMASLSAAATVPSPVSRAPVPAPPLPG--- 633
Qy 290 AKPIPTCISATGLIENRQSPATGRTPVFSPTPPPPPLPSAL-----STSSLRASMTS 346
Db 634 -----DSGTIIPPPAPGDSSTTPP-PPPPPPPPPLPGGVCISPPSLPGGTAI 682
Qy 347 TTPPP-----VPPPPP-PPATALQAP-----AVPPPPAPLQIAPGVLPAPPPIAPP 392
Db 683 SPPPLSGDATIIPPPPLPEGVGPSLPGGTAIAPPPP-PL-----PGSARIPPPP---PP 736
Qy 393 L-----VQSPPPVARAAVCETVPVHPLPQGEVQGLPPPPPP-----PPLPPPGIRPSSPV 443
Db 737 LPGSAGIPIPPP-----PLP-GEA-GMPPPPPPLPGGCGIPIPPPPFPGCP- 779
Qy 444 TVTALAHPPSGLHPTPTAPGPHVPLMP-----SPPSOVIP-ASEPKRHPSTLPLVISDAR 498
Db 780 -----GIPPPP-----PGMGMPPPPPFGVGAAPVPLPGLTPKK----- 814
Qy 499 SVLLEAIRKGLQLRK-----VEEQREQE-----AKHERIEND 530
Db 815 -----LYKPEVOLRRPNWSKLVAEDLSQDCFVTKVKEDRFENN 852

RESULT 2
US-09-080-897-4

; Sequence 4, Application US/09080897
; Patent No. 5985574

; GENERAL INFORMATION:

; APPLICANT: King, Mary-Claire

; APPLICANT: Lynch, Eric D.

; APPLICANT: Lee, Ming

; APPLICANT: Morrow, Jan E.

; APPLICANT: Welch, Piri L.

; APPLICANT: Leon, Pedro E.

; TITLE OF INVENTION: Modulators of Actin

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/080,897

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-080-897-4

Query Match 10.7%; Score 318.5; DB 2; Length 1255;
Best Local Similarity 24.2%; Pred. No. 1.7e-15;
Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;

Qy 51 IFGELFNEAHFSFRVNSLQERVDRLSVSVTOLDPKEEELSLODITMRKAFFRSSTIQDQ 110
Db 308 LINALITPAEELDFRVHI---RSELMRLGLHQLVLOELREIENEDMKVOLCVFDEQ-GDED 363
Qy 111 LFRK-----TLPIQLQETVDCVQPPPLNLTLP---YRDDGKGLKPY- 151
Db 364 FFDLKGRLDIRMDDGEVQIILNTVSKAEPHFLSLOHLLVRNDYEAPQYK 423
Qy 152 -----TNPYS---FFDLWKMKLQDTEKRR-EKRRKQKQKNDLRHPEK 192
Db 424 LIECVSQVLHKNAGADPDFCRHLQIDIERLVQIMDKTKVKESEAKATELEKKLDEL 483
Qy 193 VPRAPHRREWOKL-----ACGPELAEDDANL-----LHKHTEVANGPASH 234
Db 484 TAR---HELQVEMKKHENDFEOKLQDGEKDALDSEKQOITAAQKQDLEAEVSKLTGEVAK 541
Qy 235 PETRPQTVVDMHDSYLSALPFSOMSELLTRA-----EERVLVRPHEPPPPPPMHGAGD 289
Db 542 LSKELEDAKNEMASLSAAVAVPVSAAVAPPPLPGDSGTIVPP-PPPPPLPG--- 596
Qy 290 AKPIPTCISATGLIENRQSPATGRTPVFSPTPP-----PPPPPLSALSTSSLRASMT 345
Db 597 -----GVV---PPSP-----PLPGTCIPPPPLPGG-----A 621
Qy 346 STPPPP-----VPPPPPPATALQAPVPPPPAPLQIAPGVLPAPPPPTAPPLVOPSP 398
Db 622 CIPPPQLPGSAAIIPPPPLPGVA---SIPPPP-PLPGATAI---PPPPPLPGATAI PPP 674
Qy 399 PVARAAPVCETVPVHPLPQGEVQGLPPPPPPPP-----LPPPGIRPSPVTVTALAHPSPG 454
Db 675 P-----PLPGG--TGIPPPPPPLPGSVGVPPPPPLPGG-----G 707
Qy 455 LHPTPTAPGPHVPLMPSPSPSVIPASEP-----KRIPSTLPLVISDARSVLLEAIRGQ 510
Db 708 LPPPPPPFPFG--AFGIPPPPPMGVPPPPPPFGVGAAPVPLPGLTPKKV-----YKPEVQ 761
Qy 511 LRK-----VEEQREQE-----AKHERIEND 530
Db 762 LRRPNWSKFVAEDLSQDCFVTKVKEDRFENN 792

RESULT 3

US-08-899-595-1

; Sequence 1, Application US/08899595

; Patent No. 611072

; GENERAL INFORMATION:

; APPLICANT: Narumiya, Shuh

; APPLICANT: Takahashi, No. 6111072uaki

; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE

; TITLE OF INVENTION: ENCODING SAME

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,595
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-242701
FILING DATE: 26-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 9-90170
FILING DATE: 25-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stephen A. Bent
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 049441/0112
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-595-1

Query Match 10.7%; Score 318.5; DB 4; Length 1255;
Best Local Similarity 24.2%; Pred. No. 1.7e-15;
Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;

QY 51 IFGELNEAHSFVRNSLQVRDLRSVSTQDPKEEELSLODITMKAFRSSTIDQQ 110
DB 308 LINALITPAEELDFRVHI---RSELMRLGLHQLVQLREIENEDMKVQLCVFDEQ-GDED 363
QY 111 LFDK-----TLPIQETDVCEQPPPLNLTTP---YRDGKGLKFY- 151
DB 364 FFDLGRLLDIRMEMDDFGEVFOIILNTVKDKSAEPHFLSILQHLVLRNDYEARPOYYK 423
QY 152 -----TNPSY---FFDLWKEKMLQDTEKRR-EKRRKQKQKNDLRPEPEK 192
DB 424 LIECVSQIVLHNGTDPDFKCRHLQIDIERLVQDMIDTKVEKSEAKATELEKLDSEL 483
QY 193 VPRAPDRREWKL-----AGPELAEDANL-----LHKHIEVANGPASH 234
DB 484 TAR--HELQVEMKMEMDEFQKQLQDQGEKDALDSEKQITAQKQDLAEVSLITGEVAK 541
QY 235 FETRPOTYVDHMDGYSLSALPFSQMSSELLTRA-----EERVLRPHPEPPPPMHGAGD 289
DB 542 LSKELEDAKNEMASLAVVAVPSVSSAAVPPAPPLPGDSGTWPP--PPPPPLPG--- 596
QY 290 AKPIPTCISSATGLIENRPOSATGRTPVEVSPTTP-----PPPPPLPSALSTSSLRASMT 345
DB 597 -----GWV---PSP-----PLPGTCIPPPPLPGG-----A 621
QY 346 STPPPP-----VPPPPPPATLQAPVPPPPAPLQIAGVVLHAPPPIAPPLVQPS 398
DB 622 CIPPPPLPGSAAIIPPPPLPGVA-----SIPPPP-PLPGATAI--PPPPPLPGATAIPPP 674
QY 399 PVARAAPVCETVPVHPLPOCEVGLPPPPPPPP-----LPPPGIRPSSPVVTALAHPPSG 454
DB 675 P-----PLPGG--TGIPPPPPPLPGSVGVPVPPPLPGG-----G 707

QY 455 LHPTSTAPGPHVPLMPPSPQVIPASEP-----KRHPSTLPVISDARSVLLLEAIRKGIQ 510
DB 708 LPPPPPPPPG--APGIPPPPPGPGVPPPPPPGFGVPAAPVLPFLGLTPKKV---YKPEVQ 761
QY 511 LRK-----VEEOREOE-----AKHERIEND 530
DB 762 LRRPNWSKFVAEDLSQDCFQWTKVKEDRFENN 792

RESULT 4

US-09-323-735-4
Sequence 4, Application US/09323735
Patent No. 6197932
GENERAL INFORMATION:
APPLICANT: King, Mary-Claire
APPLICANT: Lynch, Eric D.
APPLICANT: Lee, Ming
APPLICANT: Morrow, Jan E.
APPLICANT: Welcsh, Pirl L.
APPLICANT: Leon, Pedro E.
TITLE OF INVENTION: Modulators of Actin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/323,735
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/080,897
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UW97-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-323-735-4

Query Match 10.7%; Score 318.5; DB 4; Length 1255;
Best Local Similarity 24.2%; Pred. No. 1.7e-15;
Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;
QY 51 IFGELNEAHSFVRNSLQVRDLRSVSTQDPKEEELSLODITMKAFRSSTIDQQ 110
DB 308 LINALITPAEELDFRVHI---RSELMRLGLHQLVQLREIENEDMKVQLCVFDEQ-GDED 363
QY 111 LFDK-----TLPIQETDVCEQPPPLNLTTP---YRDGKGLKFY- 151
DB 364 FFDLGRLLDIRMEMDDFGEVFOIILNTVKDKSAEPHFLSILQHLVLRNDYEARPOYYK 423
QY 152 -----TNPSY---FFDLWKEKMLQDTEKRR-EKRRKQKQKNDLRPEPEK 192

Db 424 LIEECVSVIHLKNGTDPDFKCRHLQIDIERLVDMIDKTKVKESEAKATELEKKLDSEL 483
 Qy 193 VPRAPDRRREWOKL-----AQGPDLAEDDANL-----LKHIEVANGPASH 234
 Db 484 TAR--HELQVEMKKWENDFEQLQDLOGEKDALDSEKQOITAQODLEAEVSKLTGEVAK 541
 Qy 235 FETRPQTVDHMGOSYSLSALPFSOMSELLTRA-----EERVLVRPHEPPPPPMHAGD 289
 Db 542 LSKELEDAKNEMASLSAVVAPSVSSAAVAPPPLPGDSCTVIPP--PPPPPLPG-- 596
 Qy 290 AKPIPTCISATGLIENRPOSATGRTPVFSVTPP-----PPPPPLPSALSTSSLRMSMT 345
 Db 597 -----GW-----PPSP-----PLPPGTCIPPPPPPLPGG-----A 621
 Qy 346 STPPPP-----VPPPPPPATALQAPVAPPAPPAQIAPGLHPAPPPPIAPPLVOPSP 398
 Db 622 CIPPPQLPGSAAIPPPPLPGVA-----SIPPPP-PLPGATAI--PPPPPLPGATAIPPP 674
 Qy 399 PVARAAPVCETVPVHPLPQGEVQGLPPPPPPPP-----LPPPGIRPSSPVVTALAHPPSG 454
 Db 675 P-----PLPGG--TGIPPPPPPLPGSVGVPPPPPLPGGP-----G 707
 Qy 455 LHPTPSTAGPHVPLMPSPSPQVIPASEP-----KRHPSTLPLVSDARSVLLEAIRKGIQ 510
 Db 708 LPPPPPPPPG--APGIPPPPPGMPGVPPPPPPGFGVPAAPVLPFLGTPKKV-----YKPEVQ 761
 Qy 511 LRK-----VEEQREQE-----AKHERIEND 530
 Db 762 LRRPNWSKFEVADLSQDCFWTKVKEDRENN 792

RESULT 5

US-09-080-897-2
 : Sequence 2, Application US/09080897
 : Patent No. 5985574

GENERAL INFORMATION:

: APPLICANT: King, Mary-Claire
 : APPLICANT: Lynch, Eric D.
 : APPLICANT: Lee, Ming
 : APPLICANT: Morrow, Jan E.
 : APPLICANT: Welcsh, Piri L.
 : APPLICANT: Leon, Pedro E.
 : TITLE OF INVENTION: Modulators of Actin
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : STREET: 75 DENISE DRIVE
 : CITY: HILLSBOROUGH
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 94010

COMPUTER READABLE FORM:

: COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/080,897
 : FILING DATE:

CLASSIFICATION:

: ATTORNEY/AGENT INFORMATION:
 : NAME: OSMAN, RICHARD A
 : REGISTRATION NUMBER: 36,627
 : REFERENCE/DOCKET NUMBER: UW97-001
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (650) 343-4341
 : TELEFAX: (650) 343-4342
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1248 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein

US-09-080-897-2

Query Match 10.58; Score 313.5; DB 2; Length 1248;
 Best Local Similarity 25.28; Pred. No. 4e-15;
 Matches 146; Conservative 58; Mismatches 177; Indels 199; Gaps 34;

Qy 51 IFGSLFNEAHSFSRVNSLOQERV-----DRLSVSVTQLODPKEELS--- 91
 Db 308 LINALITPAEELDFRVHTRSELMLRGLHQLVQLDLREIENEDMRVOLNVFDEQGEYSYL 367
 Qy 92 ---LQDITMKAFKRSSTIQDQQLFDRKTLPIQOETVDVCEQPPPLNILTP---YRDDGK 145
 Db 368 KGRLLDDIRMEMDDNEVEFQ-----ILLNTVKDSKAEPHFLSLQLHLLVRNDYE 416
 Qy 146 EGLKFY-----TNPSY---FFDLWKKMKLQDTEDEKREK-KRKKOKQNLND 185
 Db 417 ARPOYKLIIEECISQIVLHKNAGDPDFKCRHLQIEIGLIDQMDIKTKVKESEAKAAELE 476
 Qy 186 RPHEPEKVPRAHDRRRERWOKLAQGPDLAEDDANLLHKKHIEVANGPASHFETRPQTYVDH 245
 Db 477 KKLDELSTAR--HELQVEMKKMESDFEQKLDLQ-----GEKDALHSE-KQOIATEK 525
 Qy 246 MDGYSLSALPFSOMSELLTRAEE-----RVLVRPHEP-----PPPPPMHAGD 289
 Db 526 QDLEAEVSQLT-GEVAKLTKELEDAKKEMASLSAAAITVPFSVPSPRAPVPPAPPLPG-- 581
 Qy 290 AKPIPTCISATGLIENRPOSATGRTPVFSVTPP-----LPPPPPLPSALSTSSLRMSMTSTPP 349
 Db 582 -----DSGTIIIPPPAPGDSSTTP-PPPPPPPPPPPLPGCTAIS-----PP 621
 Qy 350 PP-----VPPPPP-PPATALQAP-----AVPPPPAPQIAPGLHPAPPPPIAPPL-- 393
 Db 622 PPLSGDATIIPPPPLPGVGIPSPSSLPGGTAIPPPP-PL---PGSARIIPPPP--PPLPG 675
 Qy 394 ---VOPSPPVARAAPVCETVPVHPLPQGEVQGLPPPPPP-----PPLPPPGIRPSSPVTVT 446
 Db 676 SAGIPPPPP-----PLP-GEA-GMPPPPPPLPGGPIPPPPPPPPPPGPPG-- 715
 Qy 447 ALAHPPSGLHPTSTAPGPHVPLMPP-----SPPSOVIP-ASEPKRHPSTLPLVISDAKSVL 501
 Db 716 -----GIPPPP---PGMGMPPPPPPGFGVPAAPVLPFLGLTPKK----- 750
 Qy 502 LEAIRKGIQLRK-----VEEQREQE-----AKHERIEND 530
 Db 751 --LYKPEVQLRRPNWSKLVADLSQDCFWTKVKEDRENN 788

RESULT 6

US-09-323-735-2
 : Sequence 2, Application US/09323735
 : Patent No. 6197932

GENERAL INFORMATION:

: APPLICANT: King, Mary-Claire
 : APPLICANT: Lynch, Eric D.
 : APPLICANT: Lee, Ming
 : APPLICANT: Morrow, Jan E.
 : APPLICANT: Welcsh, Piri L.
 : APPLICANT: Leon, Pedro E.
 : TITLE OF INVENTION: Modulators of Actin
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 : STREET: 75 DENISE DRIVE
 : CITY: HILLSBOROUGH
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 94010

COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/323,735
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/080,897
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: OSMAN, RICHARD A
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: UW97-001
;; TELEPHONE: (650) 343-4341
;; TELEFAX: (650) 343-4342
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1248 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-323-735-2

Query Match 10.5%; Score 313.5; DB 4; Length 1248;
Best Local Similarity 25.2%; Pred. No. 4e-15;
Matches 146; Conservative 58; Mismatches 177; Indels 199; Gaps 34;

QY 51 IFGELFNEAHSFVRVNSLOERV-----DRLSVSVTQDLPKEELS-- 91
Db 308 LINALTPAEELDFRVHIRELMRLGLHVOVLDRRETNEDMRVQLNVDFQGEEDSYDL 367
QY 92 ---LQDITMKAFKRSSTIQDQLFDRKTLPIQLQTYDVCEQPPPLNLTTP---YRDDGK 145
Db 368 KGRLEDDIRMEMDDNEVFQ-----ILLNTVKDSKAEPHFLSLIQLHLLVRNDYE 416
QY 146 EGLKFY-----TNPSY---FFDLWKEKMLQDTEDEK-KRKKOKQKNLD 185
Db 417 ARPOYKLIIECTISQIVLHKNGADPDFKCRHLQIEIEGLIDQMDIDTKVKESEAKAAELE 476
QY 186 RPHEPEKVPKAPDRRREWKLAQPELAEDDANLLHKHTEVANGPASHFETRPQTYVDH 245
Db 477 KKLDSLTAR--HELQVEMKKMESDFEQKLDLQ-----GEKDALHSE-KQIATEK 525
QY 246 MDGYSISALPFSQMSSELLTRAEE-----RVLVRPHEP-----PPPPMHGAGD 289
Db 526 QDLEAEVSQLT-GEVAKLTKELEDAKEMASLSAAAITVPSVPSRAPVPPAPPLPG--- 581
QY 290 AKPIPTCISSATGLIENRPOSATGRTPVFSVTPPPPPPLPSALSTSSLRASMTSTPP 349
Db 582 -----DSGTLIPPPAPGDSSTTP-PPPPPPPPPPPLPGGTAIS-----PP 621
QY 350 PP-----VPPPPP-PPATALQAP-----AVPPPPAPLQIAPGVLHPAPPPPIAPPL-- 393
Db 622 PPLSGDATIPPPPLPEGVGIPSPSSLPGGTAIPPP-PL---PGSARIPPPP--PPLPG 675
QY 394 ---VQSPPPVARAAPVCTVPHLPQGEVGLPPPPPP-----PPLPPGIRPSPSVTVT 446
Db 676 SAGIPPPPP-----PLP-GEA-GMPPPPPLPGGPIPPPPPPFGGP----- 715
QY 447 ALAHPPSGLHPTSTAPCPHVPPLMPP-----SPPSQVIP-ASEPKRHPSTLPIVSDARSVL 501
Db 716 -----GIPPPP---PGMGMPPPPPFGVGPAAPVLPFGLTPRK----- 750
QY 502 LEAIRKGILRK-----VEBQRE-----AKHERTEND 530
Db 751 --LYKPEVQLRRPNWSKLVAEDELSDQCFWTKVKEDRFENN 788

RESULT 7
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:

;; APPLICANT: John, Mallyakal E
;; APPLICANT: Barton, Kenneth A
;; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles and Brady
;; STREET: P.O. Box 2113
;; CITY: Madison
;; STATE: WI
;; COUNTRY: USA
;; ZIP: 53701-2113
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/217,327
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/812,233
;; FILING DATE: 19-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seay, Nicholas J
;; REGISTRATION NUMBER: 27,386
;; REFERENCE/DOCKET NUMBER: 1122990831
;; TELEPHONE: 608-251-5000
;; TELEFAX: 608-251-9166
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 214 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-217-327-4

Query Match 8.8%; Score 261; DB 1; Length 214;
Best Local Similarity 35.5%; Pred. No. 3.4e-12;
Matches 71; Conservative 16; Mismatches 57; Indels 56; Gaps 14;

QY 296 CISSATGLIENRPOSATGRTPVFSVTPPP---PPP-----PPLPSALSTSSLRASMTSTP 348
Db 15 CIAVA-GVLGQAPSNPPTS-Tp--ATPTPPASTTPPTTQADPTP-----TATP 58
QY 349 PPPVPPPPP---PPATALQAPAVPPPPAPLQIAPGVLHPAPPPPIAPPLVQSPPPVARAAP 405
Db 59 PVPSTPPTSSPPPVATASPPVSTPPPS-----SPPPATPP---PASPPPATPP 104
QY 406 VCETVPVHPLPQGEVQGLPPPPPPPPPLPPPPGIRPSPSVTVTALAHPPPSGLHPTSTAPG- 464
Db 105 PASPPPATPPPASPPPATPPPATPPPATPP---PATP-----PPAPLASPATVPAI 153
QY 465 --PHVPL-MPPSPSQVIPA 481
Db 154 SPVQTPLTSPAPPTA-PA 172

RESULT 8
US-08-217-327-6
; Sequence 6, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
;; APPLICANT: John, Mallyakal E
;; APPLICANT: Barton, Kenneth A
;; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles and Brady
;; STREET: P.O. Box 2113
;; CITY: Madison

```
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-217-327-6

Query Match      8.7%  Score 259.5; DB 1; Length 306;
Best Local Similarity 30.6%; Pred. No. 6.8e-12;
Matches 83; Conservative 12; Mismatches 87; Indels 89; Gaps 16;

QY 276 HEPP-----PPPMHGAGDAKPIPTCISSATGLIENRQSPATGRTPV--FVSPPTP- 325
DB 51 HSPPPHYHESPPPKH-----SPPPPTVYK-----YKSPPPMHSPPPPYHFESPPPK 101
QY 326 -PPPP-----PLPSALSTSLRASMTSTPPPPV-----PPPP----- 356
DB 102 HSPPPPTVYKYKSPPPPKHSPAPVHHYKYKSPPPPTVYKYKSPPPPKHSPAPHHYKY 161
QY 357 -PPPPALQAP-----AVPPPPALQIAPGLVHPAPPPPIAPPLVQSPPPVAAAPV 406
DB 162 KSPPPKHPPAPHHYKYKSPPPPTV-----YKKSPPPTVYKYKSPPPPKHSP- 215
QY 407 CETVPVH-----PLPQGEVGLPPPP-----PPPL-----PPPGIRPSSPVTVTALA 449
DB 216 ---APVHHYKSPPPPTVYKSPPPPHSPPPPTVYKYKSPPPMHSPPPTVYKYK 272
QY 450 HPPSGLH--PTPSTAGP---HVPLMPSPSP 475
DB 273 SPPPPMHSPPPPVYSPPPPKHHYSYTSPPPP 303

RESULT 9
US-08-580-545B-6
; Sequence 6, Application US/08580545B
; Patent No. 5932/13
; GENERAL INFORMATION:
; APPLICANT: Yoshihisa, Kasukabe
; APPLICANT: Koichi, Fujisawa
; APPLICANT: Susumu, Nishiguchi
; APPLICANT: Yoshihiko, Maekawa
; APPLICANT: Randy, Allen
; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 601 Thirteenth Street, NW
; CITY: Washington
; STATE: DC

; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,545B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bretschneider, Barry E.
; REGISTRATION NUMBER: 28,055
; REFERENCE/DOCKET NUMBER: 04473/068001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/783-5070
; TELEFAX: 202/783-2331
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-580-545B-6

Query Match      8.3%  Score 246.5; DB 2; Length 297;
Best Local Similarity 34.5%; Pred. No. 5.9e-11;
Matches 76; Conservative 8; Mismatches 91; Indels 45; Gaps 12;

QY 273 VRPHE-PPPPPMHGAGDAKPIPTCISSATGLIENRQSPATGRTPVFSPTPPPPPL 331
DB 56 VPPKAPTAPPTGPTPKP-----PTKAPTPTYKPTKAPTPTYK 97
QY 332 PSALSTSLRASMTSTPPPV-PPPPPPATALQAPVPPPPALQIAPGLVHPAPPPPIA 390
DB 98 PPA-----PAPPTKAPTPTYKPPAPAPPTKAPTPTYKPPAPAPPTKAP-----TPPFK 145
QY 391 PPLVQSPPPVAAAPVCTVPHLPQGEVGLPP--PPPPPLPPGIRPSSPVTV-TA 447
DB 146 PP--APAPT--KAPTPTYKPTTAPAPVPPKAPTPTYKPTTAPPTTKAPTAPAPTPT 201
QY 448 LAHPSPGLHPT-PSTAPGPHVPLMPPSPSOVIVASEPKR 486
DB 202 PYKPPVPTPYKPTTAP--PYKPPSP--LPPVPTTK 236

RESULT 10
US-09-262-653A-6
; Sequence 6, Application US/09262653A
; Patent No. 6166294
; GENERAL INFORMATION:
; APPLICANT: Yoshihisa, Kasukabe
; APPLICANT: Koichi, Fujisawa
; APPLICANT: Susumu, Nishiguchi
; APPLICANT: Yoshihiko, Maekawa
; APPLICANT: Randy, Allen
; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 601 Thirteenth Street, NW
; CITY: Washington
; STATE: DC

; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```


Query Match 8.3%; Score 246.5; DB 4; Length 1185;
Best Local Similarity 25.2%; Pred No. 3; le-10;
Matches 103; Conservative 51; Mismatches 149; Indels 105;

198 HDRRERWKLACGP ELAEDDANLLKHKIEVANGPASHFEETRQTVYDHMDGGSYLSA
480 HHHHHQQQQQQQQQQQQQQHGN-----SGPP-----PCAFPHPLGGSSHHH

b 526 AMSPSLGS-----LRPYPPGPA
y 313 TGRTPVEVSPTP-----PPPPPP

[illegible]

RESULT 12
US-08-642-255-32
: Sequence 32, Application US/08642255
: Patent No. 5773249

ent No. 5773249

GENERAL INFORMATION:

APPLICANT: CAPPELLO, Joseph

APPLICANT: FERRARI, Franco A.

TITLE OF INVENTION: High Molecular Weight Collagen-Like

TITLE OF INVENTION: Protein Polymers

NUMBER OF SEQUENCES: 135

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLUHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,255

AFFIDAVIT NUMBER: 05/06/0424,233
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: ROWLAND, Bertam I.
 REGISTRATION NUMBER: 20,015
 REFERENCE/DOCKET NUMBER: A55556-3/BIR
 TELECOMMUNICATION INFORMATION:

TOPOLOGY: Linear

SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids

; LENGTH: 4

;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-475-411A-65

Query Match 8.2%; Score 244; DB 4; Length 408;
Best Local Similarity 34.1%; Pred. No. 1.3e-10;
Matches 74; Conservative 5; Mismatches 92; Indels 46; Gaps 9;

QY 275 PHEPPPPPMHAGDAKPIPTCISATGLIENRQSPATGRFTVFSPTPPPPPLPSA 334
DB 51 PGPPGPPGPPGPPGAPV-----GSPGAP-----GPPGPPGPPGPPGA 89
QY 335 LSTSSLRASMTSTPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 392
DB 90 -----PGPPGPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 135
QY 393 LVQPSPPVARAAV-----CETVPVHPLQCEVOGLPPPPPPPPPPPPPPPP 448
DB 136 --GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 135
QY 449 AHPPSGLHPT-PSTAGPHVPLMPPSPSQVIPASEP 484
DB 192 PGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGSP 228

RESULT 15

US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249

; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-029A-65

Query Match 8.2%; Score 244; DB 4; Length 408;
Best Local Similarity 34.1%; Pred. No. 1.3e-10;
Matches 74; Conservative 5; Mismatches 92; Indels 46; Gaps 9;

QY 275 PHEPPPPPMHAGDAKPIPTCISATGLIENRQSPATGRFTVFSPTPPPPPLPSA 334
DB 51 PGPPGPPGPPGPPGAPV-----GSPGAP-----GPPGPPGPPGPPGA 89
QY 335 LSTSSLRASMTSTPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 392
DB 90 -----PGPPGPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 135
QY 393 LVQPSPPVARAAV-----CETVPVHPLQCEVOGLPPPPPPPPPPPPPPPP 448
DB 136 --GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 135
QY 449 AHPPSGLHPT-PSTAGPHVPLMPPSPSQVIPASEP 484
DB 192 PGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGSP 228

Search completed: June 20, 2001, 12:07:08
Job time: 23 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:06:46 ; Search time 22.61 Seconds
(without alignments)
1883.307 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 2981
Sequence: 1 MFLVKRNPIDPRHLCHTALPR.....AVEYSDSEDDSEFDEVDWLE 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	20.1	507	2 T23959	hypothetical prote
2	363	12.2	760	2 T06291	extensin homolog T
3	362.5	12.2	1201	2 G86441	unknown protein [i
4	348.5	11.7	464	2 S22697	extensin - Volvox
5	346.5	11.6	727	2 C84334	hypothetical prote
6	334	11.2	907	2 E96636	hypothetical prote
7	332.5	11.2	505	2 S72273	actin-depolymerizi
8	326.5	11.0	574	2 T38919	wiskott-aldrich sy
9	323.5	10.9	574	2 T43556	wiskott-aldrich sy
10	323.5	10.9	1188	2 S49315	extensin-like prot
11	320	10.7	1006	2 G86292	hypothetical prote
12	318.5	10.7	1255	2 T31065	diaphanous protein
13	316	10.6	699	2 T05225	extensin homolog F
14	310	10.4	744	2 E86255	hypothetical prote
15	309	10.4	620	2 S06733	hydroxyproline-ric
16	308.5	10.3	760	2 F86387	probable pto kinas
17	305.5	10.2	708	2 D96711	hypothetical prote
18	305.5	10.2	786	2 T01456	extensin homolog F
19	304	10.2	1006	2 T42731	atrophin-1 related
20	292.5	9.8	1611	2 T38236	hypothetical prote
21	291.5	9.8	280	2 T11671	extensin-like prot
22	290.5	9.7	847	2 F96531	hypothetical prote
23	287.5	9.6	439	2 S51939	chitinase (EC 3.2.
24	286.5	9.6	426	2 JQ1696	pistil extensin-11
25	286.5	9.6	1206	2 S24407	formin isoform IV
26	286	9.6	376	2 S71558	probable cell wall
27	285.5	9.6	393	2 P00479	pistil extensin-11
28	285.5	9.6	1585	2 T31611	hypothetical prote
29	285.5	9.6	5262	2 T03454	ALR protein - huma

30	285	9.6	599	2 T10798	pherophorin-s - Vo
31	284.5	9.5	666	2 B70803	hypothetical prote
32	283	9.5	369	2 S20500	hydroxyproline-ric
33	283	9.5	461	2 T10741	extensin-like prot
34	282.5	9.5	710	2 D96728	hypothetical prote
35	281.5	9.4	403	2 S52796	prpL2 protein - hu
36	280.5	9.4	416	2 JU0465	extensin precursor
37	280	9.4	502	2 A55197	wiskott-aldrich sy
38	279.5	9.4	322	2 S25299	extensin precursor
39	279.5	9.4	1058	2 T13286	capuccino gene pr
40	279	9.4	489	2 T11622	extensin class 1 p
41	278.5	9.3	362	2 A44083	meq protein - Mare
42	278.5	9.3	990	2 T14756	hypothetical prote
43	277.5	9.3	443	2 G86388	unknown protein [i
44	277.5	9.3	491	2 T07598	proline-rich prote
45	277	9.3	379	2 T05441	proline-rich prote

ALIGNMENTS

RESULT 1
T23959
hypothetical protein R06C1.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23959
R:Kershaw, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19824
A:Accession: T23959
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-507 <WIL>
A:Cross-references: EMBL:Z81106; PIDN:CAB03220.1; GSPDB:GN00019; CESP:R06C1.3
A:Experimental source: clone R06C1
C:Genetics:
A:Gene: CESP:R06C1.3
A:Map position: 1
A:Introns: 46/2; 247/3; 356/3; 421/3

Query Match	20.1%	Score	598;	DB	2;	Length	507;
Best Local Similarity	28.8%	Pred. No.	8.4e-25;				
Matches	170;	Conservative	88;	Mismatches	214;	Indels	118;
Gaps	18;						
QY	1	MFLVKRNPIDPRHLCHTALPRGI-KNELECVNTISLANIRQLSSLSKYAEDIFGELFNEA	59				
DB	1	MPLTKRAVSPVNLISRGTPSTIHRDELQCTANGTIANLVRLQLSSLSKHAENIFGEIYHDA	60				
QY	60	HSFSFRVNSLQERVDRLSVSVTQLDPKEEELSQDITMRKAFRSTTQDQOLFDRKTLPI	119				
DB	61	MIINHKSNTLQQRIDRLHKKVEDLDSNSDQATLEANMRKAFKSSMLVDQHILDRSLTPT	120				
QY	120	PLQETDVCQOPPLNTLTPTRDDGKEGLKFPYNPSYFFDLWKEMKLODTEDEKRRKQ	179				
DB	121	ALTEIYAKCDPPDLDALNEYSRSEIPALSILYTPSPFFDLWKETLKEVAERPR----	175				
QY	180	KQNLDRPHEPEKVPKPRAPHRREWOKLAQPELAEDDANLLHKKHIEVANGPASHFETRP	239				
DB	176	-----RVKSPNGSGSKPKRRKQ-----PGGGP-----	201				
QY	240	QTYVDHMDGYSLSALPFSQMSSELLTRAERVLVRPHEPPPPPMHMGAGDAKPIPTCISS	299				
DB	202	TAYNDQMHRNRQISGRINQONEVFSFPEE-----YQAP-----	236				
QY	300	ATGLIEN-RQSPATGRTPTVFVSPTPPPPPLPSAL---STSLRASMTSTPPPPVPPP	355				
DB	237	ALGLQLNFKNQHPNSMVAIPGTMHHHPQNVHPVPSQOQGAPAAARGSPNVKRPTEAPPP	296				
QY	356	-----PPPPATALQA-----PAPVPPPPAPLQIAPGVLUH---PAPPPAPPLVQSPPPVARA	403				
DB	297	VNLHDLPPPPDMSTLSDDDDDLLPPPPPPPLLMNTSIYHQLPAEPSTIQVFPSAAPPTN	356				

Db 632 PPPVHSGPPPV-----YSPPPPVYSPPPPVKSPPPPVYSPPLLPKMSPPPTQTPVNS 687

QY 483 EPKRHPs 489
| | |
Db 688 PPRPTPS 694

RESULT 6

E96636
hypothetical protein T7Pl.21 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2000
C:Accession: E96636
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khay
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Ma
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis
A:Reference number: A86141; MUID:21016719
A:Accession: E96636
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-907 <SP0>
A:Cross-references: GB:AE005173; NID:6751696; PIDN:AAF27679.1; GSPDB:GN00
C:Genetics:
A:Gene: T7Pl.21
A:Map position: 1

Query Match	11.28;	Score 334;	DB 2;	Length 907;
Best Local Similarity	25.7%;	Pred. No. 1.4e-10;		
Matches 132;	Conservative 56;	Mismatches 212;	Indels 114;	Gaps 23;
Qy	70	QERVRLSVSVTQDLPKKEELSLODITMRKAFRSSTIOQOOLFDRKTLPIPLQETVYDCE	129	
Db	222	KERFDM-----EIDEEREKKESTSPQCKTSSSRVLSPSFSDSKSFGSRNSF--CG	270	
Qy	130	QPPPLNHLTYRDCKEGL--KFTYNPSYFFDLWKEM-----LODTEDKRKEKKQKQK	182	
Db	275	SPTTPRSVLPESMGSGPRGVGFANSASHL--LNNRVOALEKLSPIDVKRLAIHLSQK	332	
Qy	193	NLDPRHEPE-----KVPRAPHDRRREWKQLAGPELAEADDANLLHKHIEVANG-----P	231	
Db	333	EAQENESNEDVIVSVEBEIKQKKEIDESIDVKME--TEESVNLDESV--VLNGEODTIMK	390	
Qy	232	ASHFETRPOTVYDHMDGYSLSALPFSQMSSELLTRAEBERVLVRPHEPPPPPPMHGAGDAK	291	
Db	391	ISSLESTSEKLNHSE-----KYENSSQLF-----PPPPPPPP-----424		
Qy	292	PIPTCISASTGLIENRPOSATGRTPVFSPPP-----PPPPPLPSAL	335	
Db	425	PPLSPIKTASLPLSPPPPTPIADIAISMPPPPPPPPPPPPVMPKLKHEAPPPPPLPPEAV	484	
Qy	336	-----STSSLRASMTSTPPPP-----VPPPPPPPATALQAAVAPPPAP	374	
Db	485	MPLKHFAPPPPTPPAFKPLKGSAPPPPPPLPTTIAAPPPPPPPRAAVAPPPPP--542		
Qy	375	LQIAPGVLHPAPPPTAPPLVQSPSVARAAPVCETVP--VHPLPQCEVOGLPPPPPPPLP	433	
Db	543	-----PGTAAAPPPPPPPCTQAAPPPPPPPPMONRAPPPPMPMGNSGSGPPPPPPMP	598	
Qy	434	-PPGIRPSSPVTVTALAHPPSGHLPTSTAPGPHVPLMPPSPPSQVPIPASEKPRHPSTLP	492	
Db	599	LANGATPPPPPMAMANGAAGPPPPPPRMGMANGAAGPPPPPP--AARSLRPKKAATKL-	656	
Qy	493	VISDARSVL---LEAIRGIGIQLRKVEEQEQAQ	523	
Db	657	-----KRSTQGLNLYRIKQ---KV-EGRDPNAK	681	

Best Local Similarity 26.4%; Pred. No. 3.1e-10; Mismatches 176; Indels 103; Gaps 21; Matches 121; Conservative 58;

QY 172 KRKEKRRKQKOKNLDRLRPHDRREKWKLAQPELAED-DANLLHKHIEVANG 230
D 144 KGSRRHAPNNSNIQPPSAAPVPP--GKENYNAGVSKSPNEPELLNSLDPSLIDSLMKMGIS 202
QY 231 PASHFETRP--QTVVDHMDGYSLSALPFSOMSELLTRAEVR--LVRRPHEPPPPPPMHG 286
D 203 QDQIAENADFVKAYLINESAGTPTSTAPPISPPSRPERKPVSVSAPAPPPPPP 259
QY 287 AGDAKPIPTCISSATGLIENRQSPATGRTVPFVSP---TPPPPPPLPSALSTSLRA 342
D 260 -----SNGTVSSPPNSPPPIAPVSMNPAINSTSKPLPPSSRSVAAALAA 306
QY 343 SMTSTPPPPVP-----PPPPPPATLAQAPAVP-----PPAPPL 375
D 307 NKRRPPPPPPPPRRNRKPPPIGNGSSNSLPPPPPPRNSNAAGSIPLPPQGRSAPP 366
QY 376 QIAPGVHLHAPPPIA-----PPLVQSPVVARAAVCETVPVHPLPQGEVQGLPPPPP 428
D 367 RSAPST--GROPPPLSSRAVSNPPAPPAIP--GRSAPA-----LPPLGNASRTSTPPVPT 419
QY 429 PPPLP-----PPGIRPSP--VVTVALHPPSLGLHTPTSTAP--GPHVPLMPSPSPVIPA 481
D 420 PPSLPPSAPSLPPSPSLPMGAPAAP--LPPSAPIAPPPLPAGMPAAPPLPPAAPAP 477
QY 482 SEPRKHSTLPVIS-----DARSVILLEIRKG-----TQLRKVEQRE----- 519
D 478 PAPAPAPAA--PVASIAELPOQDGRANLMASIRASGGMDLKSRRKVSASPSVASTKTSNPP 536
QY 520 -OEAKHERIENDVATILSR--LAVEYSDESDSEFD 553
D 537 VEAPSPNNLDALASALNQRTKVAQSDDEDEDDWD 574

RESULT 10
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C>Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 10.9%; Score 323.5; DB 2; Length 1188;
Best Local Similarity 26.6%; Pred. No. 7e-10;
Matches 118; Conservative 40; Mismatches 158; Indels 127; Gaps 23;

QY 118 PIPLOE---TYDVCQPPPLNLTPIYRDDGKGLKFTYNPSYFFDLWKEKMLQDTEKRR 174
D 770 PAPLSSPPAPQVKSSPPPVQVSP-----PPA 797
QY 175 EKRKQKOKNLDRLRPHDRREKWKLAQPELAEDDANLLHKHIEVANGPA 232
D 798 PKSPPLAPVSSPQVEKTSPPAP-----LSRPPPLAKSS---PPHVVSPP 844
QY 233 SHFETRPQTVVDHMDGYSLSALPFSOMSELLTRAEERVLVRPHEPPPP-----PPMGA 287
D 845 VVKSSPPAPV-----SSPPLTPKASPPAHVSSPPE---VVKSTPPAPTIVISPSEP 897
QY 288 GDAPKIPFCI-----SSATGLIENRQSPATGRTVPFVSPPT-----PP 327
D 898 SSPPTPVSLPPVVKSSPPPAWSSPPMTKSSPPPVVSSPPPTKSSPPPPAPVSSPP 957

QY 328 -----PPPLPSALSTSSLRASMTSTP---PPVP---PPPPPPATLAQAPAV---PPPPAP 374
D 958 ATKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKSPSPAPMSPPPPPEVKSSPPAP 1017
QY 375 LQIAPGVHLHAPPPI-----APPLVQSPVVARAAVCETVP--VHPLPQGEVQGLPP-- 425
D 1018 VSSPPPVKSSPPPPAPVSSPPPPVKSPPP---APVSSPPPVKSPPPAPISSPPPV 1073
QY 426 ---PPPPPL---PPGIRPSPSVTVALAHPPSLGLHTPTSTAP---GPHVPLMPD----- 472
D 1074 KSPPPAPVSSPPPPVK---SPPPAPVSSPPPIKSPPPAPVSSPPPPAPVKKPSPSLPPA 1131
QY 473 ---SPSQVIPASSEPKRHPSTLP 492
D 1132 PVSSPPPVVTPA--PPKKEQSLP 1153

RESULT 11
G86292
hypothetical protein AAF82153.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86292
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G86292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1006 <STO>
A:Cross-references: GB:AE005172; NID:g9927662; PIDN:AAF82153.1; GSPDB:GN00141
C:Genetics:
C:Map position: 1

Query Match 10.7%; Score 320; DB 2; Length 1006;
Best Local Similarity 28.1%; Pred. No. 8.9e-10;
Matches 120; Conservative 38; Mismatches 153; Indels 116; Gaps 22;

QY 113 DRKTLPIPL-----QETVYDVCQPPPLNLTPIYRDD---GKEGLKFTYNPSYFFD-- 159
D 198 DQTTQPPPLMLPPPPGDETPPVFSLPPLDEFPMPPIITWLPPDPVPAQTSSAEAFDQI 257
QY 160 ---LWKEKMLQDTEKREKRRKQKOKNLDRLRPHDRREKWKLAQPELAED 216
D 258 PPLVTITEATENPHNSHRH--RDENKKGDLDRN-----RRVKSRRRSR----- 300
QY 217 DANLLHKHIEVANGPASHFETRPQTVVDHMDGYSLSALPFSOMSE-----L 263
D 301 -----NCEA--FSTRCDVFRFCIFGTCGWNFPIDPCQNPFLPPPLPPPL 346
QY 264 LTRAERVLVRPHEPPPPPPPMHAGADAKPIPTCISSATGLIENRQSPATGRTVPFVSPT 323
D 347 PLPPPSLPVTPCSPPPPIIVNGAPPVPCVTCQVS-----PPP---TPV---PC 392
QY 324 PPPPPPLPSALSTSSLRASMTSTTPPPPPPPPPPP---ATALQAPVPPPPAP---LQI 377
D 393 SPPPPPTP-----VPCPPPPSPPPPPPPQPCITCVTAPA--PPPPQPCITCVI 439
QY 378 APGVHLHAP-----PPIAPLVQSPVVARAAVCETVPVHPLPQGEVQGLPPPPP 428
D 440 APASPPPOPCITCVAAPEPPPPPPQPCITCIPAPASPPPPVPPVPPVPTD---IFILPPPPP 496
QY 429 P-PPLPPPGIRPSPSVTVALAHPPSLGLHTPT---STAPGPHVPL-MPPSPSPVASEP 484

Db 497 LFPVLPSPSVTPVLPLP-----PPSAPLPPPLSSSLPPLPLVLSPPPP-----LPGGTV 549
 Qy 485 KRHPSTL 491
 Db 550 SQPPFTM 556
 RESULT 12
 T31065
 diaphanous protein homolog pl40mdia - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31065
 R:Watanabe, N.; Madaule, P.; Reid, T.; Ishizaki, T.; Watanabe, G.; Kakizuka, A.; Saito, EMO J. 16, 3044-3056, 1997
 A:Title: Pl40mdia, a mammalian homolog of Drosophila diaphanous, is a target protein for
 A:Reference number: Z20961; MUID:97357293
 A:Accession: T31065
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1255 <WAT>
 A:Cross-references: EMBL:U96963; NID:92114472; PID:92114473; PIDN:AAC53280.1
 A:Note: binds to GTP-bound form of Rho and binds to profilin
 Query Match 10.7%; Score 318.5; DB 2; Length 1255;
 Best Local Similarity 24.28; Pred. No. 1.4e-09;
 Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;
 Qy 51 IFGELENAHFSFRVNSLQERVDRSLVSTQLDPKKEELSLODITMRKAFRSSTIQDQ 110
 Db 308 LINALITPAEELDFRVH---RSELMRLGLHQLVQLREIENEDMKVQLCVFDEQ-GDED 363
 Qy 111 LFDRK-----TLPIPLQTYDVCEOPPLNLTLP---YRDGKGLKPY- 151
 Db 364 FFDLGRLLDDIRMEDDGEVFQIILNTKDSKAPFELSLQHLVLRNDYEARPOYK 423
 Qy 152 -----TNPST---FIDLKMKMLQDTEKRR-EKRRKOKNLDLRHEPEK 192
 Db 424 LIEECVSQIVLHKNGTDFKCRHLQIDIRLVDMIDKTKVSEAKATELEKLDSEL 483
 Qy 193 VPRAPDRRRRQKL-----AQPELAEDDNL-----LKHIEVANGPASH 234
 Db 484 TAR--HELQVEMKMKENDFEQKLDQLOGEKDALDSEKQITQAQKDLAEVSKLTGEVAK 541
 Qy 235 FETRPQTYVDHMDGYSLSALPFSOMSELLTRA-----EERVLPHPPEPPPMHGAGD 289
 Db 542 LSKELEDKAKNEMASLVAVVAPSVSAAVPPAPPLPGDSGTIVPP--PPPPPLPG--- 596
 Qy 290 AKPIPTCISATGLTENRPSQATGRTVPFVSPTPP-----PPPPPLPSALSTSLRASMT 345
 Db 597 -----GVV---PPSP-----PLPPGTCI PPPPLPGG-----A 621
 Qy 346 STPPPP-----VPPPPPPATALQAPVPPPPAPLQIAPGVLPHPAPPTIAPLVPQSP 398
 Db 622 CIPPPPLPGSAAIPPPPLPGVA---SIPPPP-PLPGATAI---PPPPPLPGATAI PPP 674
 Qy 399 PVARAAPVCEVPHVPLPQGEVQGLPPPPPPPP-----LPPPGIRPSSPVTTALAHPSG 454
 Db 675 P-----PLPGG--TGIPPPPPPLPGSGVPPPPPPPLPGG-------G 707
 Qy 455 LHPTPSTAPGPHVPLMPSPSPSVPQIPASEP-----KRHPSTLPLVSDARSVLEAIRKGIQ 510
 Db 708 LPPPPPPPPG--AGCIPPPPPCMGVPPPPPPGFGVPAAPVLPFLGLTPKKV---YKPEVQ 761
 Qy 511 LRK-----VEEQREQE-----AKHERIEND 530
 Db 762 LRRPNWSFEVLEDLSQDCFTWKVREDRFPEN 792
 RESULT 13
 T05225
 extensin homolog Fl715.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T05225
 R:Bevan, M.; Vitale, D.; Liguori, R.; Argiriou, A.; De Simone, V.; Hoheisel, J.; Mewe submitted to the Protein Sequence Database, July 1998
 A:Reference number: Z15404
 A:Accession: T05225
 A:Molecule type: DNA
 A:Residues: 1-699 <BEV>
 A:Cross-references: EMBL:AL031032
 A:Experimental source: cultivar Columbia; BAC clone F1715
 C:Genetics:
 A:Map position: 4
 A:Note: F1715.160
 Query Match 10.6%; Score 316; DB 2; Length 699;
 Best Local Similarity 25.6%; Pred. No. 9.6e-10;
 Matches 116; Conservative 40; Mismatches 159; Indels 138; Gaps 18;
 Qy 125 YDCEQPPPLNLTIPYRDDGKGLKFTYNPSYFFDLNKKMLQDTE---KRKKRKOKQ 181
 Db 342 HNICOLPNLNLTYSYNYFGGGGSCVPGGSR-----KEIALDDTRNCLASRPEORSAGE 396
 Qy 182 --KNLDRP-----HEPEKVPRAPHDRRRREWKQKLAQGPDLAEDDA 218
 Db 397 CAVVINRPVDCSKDKCAGGSSTPSKPSVHKPTVPVTP----- 435
 Qy 219 NLLKHIEVANGPASHFETRPQTYVDHMDGYSLSALPFSOMSELLTRAER---VLVRP 275
 Db 436 --VHKPTVPVTPVQKPSVPTTPVQKPS---PVPTTTPVHEPSPVLPATVDKPSVPVSRP 490
 Qy 276 HEPPPPMHMGAGDAKPIPTCISATGLIENRPSQATGRTVPFVSPTPPPPPLPSAL 335
 Db 491 VOKQPP-----KESQPDPPYDQSPVTKRSPPPAPVNSPPPPVSPPPPPPVHS-- 542
 Qy 336 STSLRASMTSTPPPPPPPPPPPPPPATALQAPVPPPPAPLQIAPGVLPHPAPPTI---APP 392
 Db 543 -----PPPPVHSPPPPP---VYSPPPPPPP-----VHSPPPPVSPPPP 578
 Qy 393 LVQSPSPVARAAPVCEVPHV-PLPQGEVQGLPPPPPPPPPLPPPGIRPSSPVTTALAH 451
 Db 579 VYSPPPPVHSPPP-----PVHSPPPPPAPVHSPPPPVHSPPPPPPVYSP-----P 622
 Qy 452 PSGLHPPTAPGPHVPLMPSPSPSVPQIPASEPKRHPSTLPLVSDARSVLEAIRKGIQ 511
 Db 623 PPVSPSPSPSP-PVVYSPPPPPPKINSPP---PVQSPPPAPV----- 660
 Qy 512 RKVEEQREQEAKHERIENDVATILSRRIAVEYS 544
 Db 661 ---EKKETPPAHAPAPSDDEFIIPPFIGHQYA 689
 RESULT 14
 E86255
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86255
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E86255
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-744 <STO>

A:Cross-references: GB:AE005172; NID:g3157926; PIDN:AAC17609.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match

Best Local Similarity 10.4%; Score 310; DB 2; Length 744;

Matches 97; Conservative 26; Mismatches 113; Indels 70; Gaps 15;

QY 249 SYLSALPFSOMSELTTRAERVLRPHPP-----PPMHGAGDAKPIPTCISS--AT 301

DB 438 AYSPPPPYSKMSPSRVATPPPPSPPPPVYVSSPPPPPVYSSPPPPPVYSSPPPP 497

QY 302 GLIENRQSPATGRTP-----VFVSTPPPPPLPSALSTSS----LRASMTSTPPPPVP 353

DB 498 PVIYSSPPPPPVYVSSPPPPPVYVSSPPPPPVYVSSPPPPPVYVSSPPPPPVYV 557

QY 354 ----PP-----PPPPATALQAPAV--PPPPAPLQIATAGVLPAPP-PIAPPLVQSPPVARA 403

DB 558 VYYPVYVQSPPPSPVYVPPVNTNSPPPPSPVYVPPVYVPPVYVPPVYVPPVYV 613

QY 404 APVCERVVHPLPQGEVQGLPPPP-----PPPLP-----PPGIRPSPV--- 443

DB 614 -----PPSPLYPPVTPSPPPSPVYVPPVTPSPPPSPVYVPPVTPSPPPSPVYV 666

QY 444 TVTALAHPPSGLHPTSTAPG-----HVPLMPPS-----PPSQVIPASEPKRHPST--- 490

DB 667 SETQSPPTTEYYYSQSPPPTKACKEGHPQATPSYEPPEYSYSSPPPPSPPTSYFP 726

QY 491 -LPVIS 495

DB 727 PMPSVS 732

RESULT 15

S06733

hydroxyproline-rich glycoprotein precursor - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000

C:Accession: S06733

R:Keller, B.; Lamb, C.J.

Genes Dev. 3, 1639-1646, 1989

A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene

A:Reference number: S06733; MUID:90128263

A:Accession: S06733

A:Molecule type: DNA

A:Residues: 1-620 <KEL>

A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867

C:Superfamily: hydroxyproline-rich glycoprotein

C:Keywords: glycoprotein

Query Match

Best Local Similarity 10.4%; Score 309; DB 2; Length 620;

Matches 90; Conservative 9; Mismatches 90; Indels 70; Gaps 15;

QY 279 PPPPMHAGDAKPIPTCISSATGLIENRQSPATGTPVFSPT-----PPPP-- 327

DB 370 PPPPP-----SSPPPSFSPPTTEQSPPPPPPAYSPPLPAPPTYSPPPTYSPPPTTY 423

QY 328 --PPPLPSALSTSLRASMTSTPPPV---PPPP---PPPATALQAP---AVPPPPAPLQ 376

DB 424 AQPPLPP-----TYSPPPPAYSPPTTYSPPPTYSPPPPPAYAQPPPPPTTY 472

QY 377 IAPGLVHPAPPI-----APPLVQSPPVARAAP---VCETVPVH---PLPQGEVQGLPPP 426

DB 473 SPPPPAYSPPPSPVYVSSPPPPVQPLPTTSPPPPRRRLHPPPPHQRPPPTTYGQPPS 532

QY 427 P-----PPPL-----PPGIRPSPVTV-----TALAHPPSGLH--PTPSTAGPHV 467

DB 533 PPTSPPPPRQIHSPPPPHWQPTTPTTYGQPPSPPTTSAPPPRQIHSPPPPHQRPPPT 592

QY 468 PL--MPPSPSQVIFASEP 484

DB 593 PTYGQPPSPPTTYSPPSP 611

Search completed: June 20, 2001, 12:09:05

Job time: 139 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:06:45 ; Search time 23.2 Seconds
(without alignments)
1377.338 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 2981
Sequence: 1 MPLVKRNPRLHCHTALPR.....AVEYSDSEDDSEFDEVDWLE 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	360	12.1	787	19 W37152	Mouse neural Mena+
2	357	12.0	783	19 W37151	Mouse neural Mena+
3	356.5	12.0	802	19 W37153	Mouse neural Mena+
4	350.5	11.8	505	19 W46889	Human Neural-Wiskott
5	342.5	11.5	501	19 W46890	Rat Neural-Wiskott
6	332.5	11.2	505	19 W53808	N-Wiskott-Aldrich
7	320.5	10.8	1315	20 W76734	Human mdia Rho tar
8	318.5	10.7	1255	19 W52249	Mouse Rho target p
9	318.5	10.7	1255	20 W76733	Mouse mdia Rho tar
10	317.5	10.7	572	18 W31855	Mycobacterium tube
11	317.5	10.7	763	18 W31852	Mycobacterium tube

12	313.5	10.5	1248	20 Y13464	Human diaphanous p
13	305.5	10.2	1012	20 Y17406	Human atrophin-1 r
14	297	10.0	520	22 B49336	Murine WASP protel
15	289.5	9.7	987	21 B43064	Human ORFX ORF2828
16	289.5	9.7	1253	21 B29663	Human tyrosine pho
17	289.5	9.7	1264	21 B29664	Human tyrosine pho
18	289.5	9.7	1274	20 W89253	Human ALP. Homo s
19	289.5	9.7	1583	21 B29662	Human tyrosine pho
20	289.5	9.7	1636	21 B29661	Human histidine do
21	287.5	9.6	439	13 R28150	Sugar beet chitina
22	286	9.6	763	21 G38942	Arabidopsis thalia
23	284.5	9.5	666	19 W72911	Mycobacterium tube
24	284.5	9.5	666	20 Y21928	Amino acid sequenc
25	284.5	9.5	666	22 B19846	Mycobacterium tube
26	281.5	9.4	504	20 Y24091	Human Wiskott-Aldr
27	265	8.9	3119	19 W72204	HSV-2 strain S85 C
28	261	8.8	214	17 R86913	Cotton fiber-speci
29	260.5	8.7	428	21 G29606	Arabidopsis thalia
30	260	8.7	647	21 B53462	Human colon cancer
31	259.5	8.7	306	17 R86912	Carrot extensin.
32	257	8.6	191	21 G29836	Arabidopsis thalia
33	257	8.6	199	21 G29835	Arabidopsis thalia
34	256	8.6	541	19 W37148	Mammalian Ena (Men
35	253	8.5	641	21 Y82327	Human Npw38Bpl tra
36	253	8.5	641	21 Y67469	Np70 protein sequen
37	250.5	8.4	470	21 G49660	Arabidopsis thalia
38	247.5	8.3	280	14 R42055	Insecticidal/Pungi
39	246.5	8.3	297	18 W15761	Cotton fibrous tis
40	246.5	8.3	399	21 B43375	Human ORFX ORF3139
41	246.5	8.3	1185	20 Y33497	Human atrophin I p
42	245	8.2	900	21 B42321	Human ORFX ORF2085
43	244	8.2	330	19 W57645	Collagen-like poly
44	244	8.2	408	17 W07539	Collagen like prot
45	244	8.2	408	22 B64007	CLP protein sequen

ALIGNMENTS

RESULT 1
W37152
ID W37152 standard; Protein; 787 AA.
XX
AC W37152;
XX
DT 06-JUL-1998 (first entry)
XX
DE Mouse neural Mena++ protein.
XX
KW Neural Mena++ protein; mammalian Ena; Enabled protein; Evl protein;
KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW cell growth; cell motility; mouse.
XX
OS Mus musculus.
XX
PN W09801755-A1.
XX
PD 15-JAN-1998.
XX
PF 03-JUL-1997; 97WO-US11669.
XX
PR 05-JUL-1996; 96US-0675815.
XX
PR (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
XX WPI; 1998-101197/09.
XX
PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
PT - used in control of cytoskeletal dynamic events in normal and
PT abnormal cell morphology, adhesion, motility, growth and

Db 128 rkavtdllgrrrksekrrdapgng-nlpmatvd-----iknpeittnrfysqvn 178
QY 220 LLH-----KHIEVAN-SPASHFE-----TRPQTYVD-----HMDGSYSLSAL 255
Db 179 ishtkekkkgakkrltkadigtspnfhghvghwdpntgfdlnldpeiknldfmcgi 238
QY 256 PFSQMSSELTRAERVL-----VRPHEPPPPPMHAGADAKPIPTCISS 299
Db 239 seaqlkd-----retskviydfiektgveavknelrrqapppppsg-gppppppphss 294
QY 300 ATGLIENRQSPATGRTPVFVSPTPP-----PPPLPSALSSSLRASMTSTPPPPV 352
Db 295 gp-----ppppargrg-----apppppscaptaappppps-----rpgvvvpppppn 337
QY 353 PP-PPPPATALQAPVPPPPAPLQIAPGVLHPAPPPIAPPLVQSPPPVARAAPVCETVP 411
Db 338 rnypppppalpsapsagppppplasma-gstappppp----- 373
QY 412 VHPLPQGEVQGLPPPPPPPLPPPPGIRPSSPVTVTALAHPPSGLHPTSTAPGPHVPLMP 471
Db 374 -----pppppppppppgl-----psgdghqvpass-gnkaall- 406
QY 472 PPSPOVIPASEPKR-HPSTLPIVSDARSVLLEAKGQLKRVKEQREQ-----EAKHER 526
Db 407 -----dqiregaqlkveqnsrpscgrdallidqirgqiklksvsgqgestpbtptsg 462
QY 527 IENDVATILSRRIAVEYSDSEDDSEFDEVDW 557
Db 463 ivgalmevmqkrskaihssdededddedf 493
RESULT 6
W53808
ID W53808 standard; Protein; 505 AA.
XX
AC W53808;
XX
DT 06-JUL-1998 (first entry)
XX
DE N-Wiskott-Aldrich syndrome protein.
XX
KW N-Wiskott-Aldrich syndrome protein; N-WASP; actin polymerisation;
KW bone structure formation.
XX
OS Bos sp.
XX
PN JP10087697-A.
XX
PD 07-APR-1998.
XX
PF 11-SEP-1996; 96JP-0261187.
XX
PR 11-SEP-1996; 96JP-0261187.
XX
PA (MITS-) MITSUI SEIYAKU KOGYO KK.
PA (TAKE/) TAKENAWA T.
XX
WPI: 1998-267125/24.
DR N-PSDB; V20072.
XX
Bovine N-Wiskott-Aldrich syndrome protein - useful for controlling
PT polymerisation of actin in bone formation
XX
PS Claim 1; Page 13-14; 19pp; Japanese.
XX
CC This sequence is the bovine N-Wiskott-Aldrich syndrome protein
CC (N-WASP) of the invention. N-WASP is useful for controlling the
CC polymerisation and depolymerisation of actin during the formation of bone
CC structure.
XX
SQ Sequence 505 AA;

Query Match 11.2%; Score 332.5; DB 19; Length 505;
Best Local Similarity 23.5%; Pred. NO. 1.7e-14;
Matches 135; Conservative 66; Mismatches 145; Indels 229; Gaps 31;
QY 129 EQPPPLN-----ILTPYRDD-----GKE-----GLKFTY----- 152
Db 6 qpppprrvtnvgsllltpqeneslftflgkckvtmssavvqlyaaadrncmwmkksqva 65
QY 153 -----NP-SYF-----FD-----LMKEMLQD----- 171
Db 66 clivkdnqrsyflrfdikdgkllweqelynnfyvnsprgyfhtagdtcvalnfancee 125
QY 172 KRKERRK-----QKQKNDLRHPEKVPRAP----- 199
Db 126 eakkrkavtdilgrkrksekrirdpnpngnlpmatvdknpeittnrfygpqinnisht 185
QY 200 RRREWQKLAQGPELAEDD-----ANLLKHIEVANGPASHFETRPQTYVD-----HMDGS 249
Db 186 kekkgk-akkrlltkadigtspnfhg-----ighvwdpntgfdlnldpeiknld 235
QY 250 YLSALPFSQMSSELTRAERVL-----VRPHEPPPPPMHAGADAKPI 293
Db 236 fmcgiseaqlkd-----retskviydfiektgveavknelrrqapppppsgg----- 286
QY 294 PTCISATGLIENRQSPATGRTPVFVSPTPPPPPLPSALSTSLRASMTSTPPPPV- 352
Db 287 -----ppppppppphs-----sgpppppar 305
QY 353 --PPPPPPATALQAPVPPPPAPLQIAPGVLHPAPP--IAPPLVQSPPPVARAAPVCE 408
Db 306 grgapppppsra--ptaappppppsr--pvgapppppparmyp--pialpssap-- 355
QY 409 TVPVHPLPQGEVQG-LPPPPPPPPPLPPPPGIRPSSPVTVTALAHPPSGLHPTSTAPGPHV 467
Db 356 sgppppppplsvsgvapppppppppppg--pppppgl-----psgdghqvtpa-gska 407
QY 468 PLMPPSPSQVIPASEPKR-HPSTLPIVSDARSVLLEAKGQLKRVKEQREQ-----EA 522
Db 408 all-----dqiregaqlkveqnsrpscgrdallidqirgqiklksvtdapestppapa 462
QY 523 KHERIENDVATILSRRIAVEYSDSEDDSEFDEVDW 557
Db 463 psglvgalmevmqkrskaihssdededddedf 497
RESULT 7
W76734
ID W76734 standard; Protein; 1315 AA.
XX
AC W76734;
XX
DT 15-JAN-1999 (first entry)
XX
DE Human mDia Rho targeting protein.
XX
KW Rho protein; mDia; mammalian diaphanous; target protein; human;
KW Rho protein-combining; proline-combining; chromosome 5q31.2; disease;
KW respiratory tract; oversensitiveness; bronchial asthma; marrow leukaemia;
KW marrow dysplasia syndrome.
XX
OS Homo sapiens.
XX
PN JP10262680-A.
XX
PD 06-OCT-1998.
XX
PF 25-MAR-1997; 97JP-0090170.
XX
PR 25-MAR-1997; 97JP-0090170.
XX
PA (KIRI) KIRIN BREWERY KK.
XX

DR WPI; 1999-002481/01.
 DR N-PSDB; V62933.
 XX Human Rho target protein and its gene - useful for elucidation of
 PT mechanisms of respiratory tract disease
 XX
 PS
 PS Claim 2a; Page 28-31; 54pp; Japanese.
 XX
 CC This sequence represents a human Rho target protein, mDia (mammalian
 CC diaphanous). This protein has active type Rho protein-combining ability,
 CC has proline-combining ability, has a M.W. of 150 kDa measured by sodium
 CC dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and its
 CC gene is positioned at human chromosome 5q31.2. The nucleic acid can be
 CC used for the recombinant production of the protein. The protein is useful
 CC for the elucidation of mechanism of diseases such as respiratory tract
 CC oversensitiveness, bronchial asthma, acute marrow leukaemia and marrow
 CC dysplasia syndrome.
 XX
 SQ Sequence 1315 AA;

Query Match 10.8%; Score 320.5; DB 20; Length 1315;
 Best Local Similarity 25.4%; Pred. No. 3e-13;
 Matches 148; Conservative 59; Mismatches 183; Indels 193; Gaps 34;
 QY 51 IFGELFNEAHSFRVNSLQERV-----DRLSVSVTOLDPKEELS--- 91
 Db 360 linalitpaeeeldfrvhrselmrllghvqlqlreienedmrqinvfdegeedsydl 419
 QY 92 ---LQDITMRKAFRSTIQDQQLFDRKTLPIQLQETVDCVQPPPLNLITP---YRDGK 145
 Db 420 kgriiddirmemddnevfq-----llntvkskaephflsilqhlvlrvndye 468
 QY 146 EGLRFY-----TNPSY---FFDLWKEKMLQDTEKRR-EKRRQKQKMLD 185
 Db 469 arpgyykileescisqvlhknagdpdkfcrhlqlieqlidqmdktkvekseakaale 528
 QY 186 RPHEPEKVPRAHRRREWKQLAOGPELAEDDANLLKHIEVANGPASHFETRPQTVVDH 245
 Db 529 kklidseltar--helqvemkmesdfeqlqdlq-----gekdlhse-kqiatek 577
 QY 246 MDGYSLSALPFSOMSELLTRAEE-----RVLRPHEP-----PPPPMHGAGD 289
 Db 578 qdleavsqilt-gevaktileedakkemasalsaaaitvpsvpsrapvpappilpg--- 633
 QY 290 AKPIPTCISSATGLIENRPOSATGRTPVFVSPTPPPPPPLPSAL---STSSLRASMTS 346
 Db 634 -----dsgrtliappppagdsttpp-ppppppppppplpggvcissppslpggtai 682
 QY 347 TTPPPP-----VPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 392
 Db 683 sppppslgdatlppppplpegvgipssslpggtaiapppp-pl---pqsaripppp---pp 736
 QY 393 L-----VQSPPPVRAAPVCTVPVHPLPQGEVQGLPPPPPPPPPPPPPPPPPPPP 443
 Db 737 lpgsagippppp-----plp-gea-gmppppppplpgggaipppppfpgggp- 779
 QY 444 TVTALAHPPSLGHTPTSTAPGHVPLMPP-----SPPSOVIP-ASEPKRHPSTLPVISAR 498
 Db 780 -----gipppp-----pamgmppppppfpgvpaapvlpfgltppk----- 814
 QY 499 SVLLEAIRKGLQLRK-----VEEQEOE-----AKHRIEND 530
 Db 815 -----lykpevgllrpnwsklvaedlsqdcfwtkvkedrfenn 852

RESULT 8

W52249
 ID W52249 standard; Protein; 1255 AA.
 XX
 AC W52249;
 XX
 DT 15-JUN-1998 (first entry)

XX Mouse Rho target protein p140mDia.
 DE
 DE Rho target protein; p140mDia; mouse; Rho protein combining activity;
 KW profilin combining activity; inhibitor; actin cell structure.
 XX
 OS Mus sp.
 XX
 PN JPI0067798-A.
 XX
 PD 10-MAR-1998.
 XX
 PF 26-AUG-1996; 96JP-0242701.
 XX
 PR 26-AUG-1996; 96JP-0242701.
 XX
 PA (KIRI) KIRIN BREWERY KK.
 XX
 DR WPI; 1998-225192/20.
 DR N-PSDB; V19788.
 XX
 PT Rho target protein p140mDia - useful for, e.g. reconstituting actin
 cell structure
 XX
 PS Claim 12; Page 18-23; 28pp; Japanese.
 XX
 CC This sequence is the protein of the invention, which is the Rho
 CC target protein p140mDia (1). The protein has active type Rho protein
 CC combining activity and profilin combining activity. (1) can be used in
 CC screening for its inhibitors, by measuring the extent of its inhibition
 CC by a compound exhibiting inhibitory properties. The protein is used for
 CC reconstituting actin cell structure.
 XX
 SQ Sequence 1255 AA;
 Query Match 10.7%; Score 318.5; DB 19; Length 1255;
 Best Local Similarity 24.2%; Pred. No. 3.9e-13;
 Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;
 QY 51 IFGELFNEAHSFRVNSLQERVDRLSVSVTOLDPKEELSLODITMRKAFRSTIQDQ 110
 Db 308 linalitpaeeeldfrvhi---rselmrllghvqlqlreienedmkvqlcvfdeq-gded 363
 QY 111 LFDK-----TLPIQLQETVDCVQPPPLNLITP---YRDGKGLKRFY- 151
 Db 364 ffdlkgriiddirmemddafgevfqiilntvkskaephflsilqhlvlrvndyearpgyk 423
 QY 152 -----TNPSY---FFDLWKEKMLQDTEKRR-EKRRQKQKMLDRPHPEK 192
 Db 424 lieecvsqvlhknagtdpdkfcrhlqidierlvqmdktkvekseakatelekkldsel 483
 QY 193 VPRAPHRREWKQL-----AOGPELAEDDANL-----LKHIEVANGPASH 234
 Db 484 tar--helqvemkmesdfeqlqdlqgekdaldsekkqitqkdleavskitgevsk 541
 QY 235 FETRPQTVVDHMDGYSLSALPFSOMSELLTRA-----EERVLVRPHEPPPPPPMHGAGD 289
 Db 542 lskeledaknemalsavvvavpsvssaaavppappplpgdsgtviapp-pppppplp- 596
 QY 290 AKPIPTCISSATGLIENRPOSATGRTPVFVSPTPPP-----PPPPPLPSALSTSSLRASMT 345
 Db 597 -----gvv-----ppsp-----plpggtcippppplpgg-----a 621
 QY 346 STTPPPP-----VPP 398
 Db 622 cippppplpgsaaippppplp-gva-----sipppp-plpggtai---ppppplpggtaiapp 674
 QY 399 PVARAAPVCTVPVHPLPQGEVQGLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 454
 Db 675 p-----plpgg--tgipppppplpgsvgppppplpgg-----g 707
 QY 455 LHPTPTAPGHVPLMPPPPSPQVIPASEP-----KRHPSTLPVISDARKSVLLEAIRKGIQ 510

Db 708 lppppppfpg--apglpppppgmgvppppfpgfvgpaapvlpfgltpkvv-----ykpveq 761
 QY 511 LRK-----VEQREQE-----AKHERIEND 530
 Db 762 lrrpnwskfvaedlsqdcfkwkdrfenn 792
 RESULT 9
 ID W76733 standard; Protein; 1255 AA.
 AC W76733;
 DT 15-JAN-1999 (first entry)
 DE Mouse mbia Rho targeting protein.
 KW Rho protein; mbia; mammalian diaphanous; target protein; murine;
 KW Rho protein-combining; proline-combining; chromosome 5q31.2; disease;
 KW respiratory tract; oversensitiveness; bronchial asthma; marrow leukaemia;
 KW marrow dysplasia syndrome.
 OS Mus sp.
 PN JP10262680-A.
 PD 06-OCT-1998.
 PE 25-MAR-1997; 97JP-0090170.
 PR 25-MAR-1997; 97JP-0090170.
 PA (KIRI) KIRIN BREWERY KK.
 WP1: 1999-002481/01.
 N-PSDB; V62932.
 Human Rho target protein and its gene - useful for elucidation of
 mechanisms of respiratory tract disease
 Example 4; Page 19-22; 54pp; Japanese.
 This sequence represents a mouse Rho target protein, mbia (mammalian
 diaphanous) which is used in a method to extract a human mbia protein.
 This human protein has active type Rho protein-combining ability, has
 proline-combining ability, has a M.W. of 150 kDa measured by sodium
 dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and its
 gene is positioned at human chromosome 5q31.2. The nucleic acid can be
 used for the recombinant production of the protein. The protein is useful
 for the elucidation of mechanism of diseases such as respiratory tract
 oversensitiveness, bronchial asthma, acute marrow leukaemia and marrow
 dysplasia syndrome.
 Sequence 1255 AA;
 Query Match 10.7%; Score 318.5; DB 20; Length 1255;
 Best Local Similarity 24.2%; Pred. No. 3.9e-13;
 Matches 138; Conservative 64; Mismatches 192; Indels 177; Gaps 30;
 QY 51 IFGLFNEAFSFRVNSLQERVDRLSVYTLQDPKKEELSLODITWRKAFRSTIQDQQ 110
 Db 308 linalitpaeeeldfrvhi-----rselmlrlghqlrelenedmkvqlcvfdeq-gded 363
 QY 111 LFRK-----TLPILOETVDVCEQPPPLILTP---YRDGREGLEKFY- 151
 Db 364 ffdlkgrrldirmemddofgevfiintvkdkaephflsilqhlilvrndyearpyyk 423
 QY 152 -----TNPSY-----FFDLWKEKMLQDTEKRRK-EKKOKQKNLDRPHEPEK 192
 Db 424 lieecvqlvlnkngtdpdkrhlqldierlvdqmdlktkvekseakelekklidse 483

QY 193 VPRAPHDRREWOKL-----AQCPLEAEDDANL-----LHKHIEVANGPASH 234
 Db 484 tar--helgvemkkmendfeqlqldqgekdaldsekqitakqdeaeavskltgevak 541
 QY 235 FETRPQTYVDHMDGSYSLSALPFSQMSSELLTRA-----EERVLRPHRHEPPPPPMHAGD 289
 Db 542 lskeledaknemaslsavvvapsvssaaavppapplpqdsdgtvipp--pppppplpq--- 596
 QY 290 AKPIPTCISSATGLIENRPOSATGRTPVFSPTPP-----PPPPPLPSALSTSSLRASMT 345
 Db 597 -----gvv--ppsp-----plppgtcippppplpgg-----a 621
 QY 346 STPEPP-----VPPPPPPATALQAPVPPPPAPLQIAPGVLPAPPTIAPPLVQVSP 398
 Db 622 .cippppqlpgsaalpbbppplpgva-----sippbp-plpgatai--ppppplpataippp 674
 QY 399 PVARAAPVCETVPVHPLPQCEVOGLPPPPPPPP-----LPPPGIRPSSPVVTALAHPPSG 454
 Db 675 p-----plpgg--tgippppplpgsvgvpbbppplpggp-----g 707
 QY 455 LHPTPSTAGPHVPLMPSPSQVIPASEP-----KRHPSTLPVISDARSVLEAIRKGIQ 510
 Db 708 lppppppfpg--apglpppppgmgvppppfpgfvgpaapvlpfgltpkvv-----ykpveq 761
 QY 511 LRK-----VEQREQE-----AKHERIEND 530
 Db 762 lrrpnwskfvaedlsqdcfkwkdrfenn 792
 RESULT 10
 ID W31855 standard; Protein; 572 AA.
 AC W31855;
 DT 27-APR-1998 (first entry)
 DE Mycobacterium tuberculosis 55 kDa protein.
 KW Tuberculosis; mycobacteria; infection; diagnosis;
 KW antimycobacterial; antibiotic; vaccine.
 OS Mycobacterium tuberculosis.
 PN WO9741252-A2.
 PD 06-NOV-1997.
 PF 18-APR-1997; 97WO-EP01973.
 PR 29-APR-1996; 96DE-4017184.
 PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 PI Espitia C, Honisch C, Moreno C, Singh M;
 WP1: 1997-549750/50.
 N-PSDB; T93610.
 New DNA and related proteins or RNA derived from M. tuberculosis -
 used for diagnosis of mycobacterial infections, monitoring
 vaccination and development of anti-mycobacterial agents
 Claim 11; Fig 16; 55pp; English.
 This novel 55 kDa protein is encoded by an open reading frame of
 a Mycobacterium tuberculosis DNA fragment (see T93610) containing
 polymorphic GC-rich sequences. Its amino acid sequence shows
 a high proline content, but there is no homology to any known
 proline-rich antigens of mycobacteria. Novel M. tuberculosis
 proteins (see W31851-57) are claimed. These can be produced as
 recombinant proteins, especially in bacterial, yeast, fungal or
 higher eukaryote host cells, and used for diagnosing tuberculosis

Db 309 lqhpqppqfqlppqasqgqapltgtsaaayphtslqlipasqsalqsgqppregplppap 368
QY 402 RAAPCVTPVHPLPQ-----GEVQG-----LPPPP----- 427
Db 369 lamphikpppttpqlpqpqahkbphtsgpspsfsmnanlppppalkplsstshhpps 428
QY 428 ---PPP-----PLP-----PPGIRPSPVTVTAHAPPSGLH---PTPSTA-----PG 464
Db 429 ahppplqimpqspqlpsapqppgltqsnlppppashphtglhvaqpqfahqpfvpg 488
QY 465 PHVPLMPPSPSQVTP-----AS 482
Db 489 gpppttpctpstcpagpgtsaqpcsgaaasgsgsagsscpplptvqikeealdade 548
QY 483 EPK-----RHPSTLPVTSARSVLEAIR-----KGQLRKV 514
Db 549 epeppppprpspeptvtdtpshasqarfkykhldrgynscartdlyfmplagskiak 608
QY 515 EQREQEAKHE 525
Db 609 reeaiekakre 619

RESULT 14
B49336
ID B49336 standard; peptide; 520 AA.
XX B49336;
AC B49336;
DT 01-MAR-2001 (first entry)
DE Murine WASP protein.
KW Sam68; arginine methylation; arginine methyltransferase activity; mouse;
KW cancer; Wiskott-Aldrich syndrome; antibody; SLM-1; SLM-2; WASP; hnRNP K.
XX Mus sp.
XX CA2266760-A1.
XX 08-OCT-2000.
XX 08-APR-1999; 99CA-2266760.
XX 08-APR-1999; 99CA-2266760.
XX (DAVI-) DAVIS-JEWISH GEN HOSPITAL MORTIMER B.
PI Richard S;
XX WPI: 2001-050180/07.
XX Assaying arginine methyltransferase activity useful for determining
PT whether a cell is cancerous or has cancer potential comprises measuring
PT the methylation of specific substrates such as SLM-1, SLM-2, Sam68 -
PS Disclosure; Fig 5; 25pp; English.
XX The present invention provides a novel method for assaying arginine
CC methyltransferase activity in a cell by measuring the methylation of
CC certain proteins. These proteins are Sam68, SLM-1, SLM-2, hnRNP K and
CC WASP. Also provided are antibodies capable of recognising proteins
CC containing methylated arginines. These can be used in the stimulation of
CC an immune response and to measure the degree of methylation of the
CC proteins. The method is useful in the diagnosis of cancer, cancerous
CC cells and Wiskott-Aldrich syndrome. It is also useful for measuring the
CC extent to which ligands are binding to receptors.
XX Sequence 520 AA;

Query Match 10.0%; Score 297; DB 22; Length 520;
Best Local Similarity 24.3%; Pred. No. 3.9e-12;

Matches 133; Conservative 36; Mismatches 138; Indels 240; Gaps 27;
QY 122 QETVDVCQPPPLNLTTPY-----RDGKEGLKF--YTNPSYFFDLWKEKMLQDTEKRRK 174
Db 101 qelysqlyvltp-----tpffhtfagddcvglnfadeseaqfalvqeki-----qkrn 151
QY 175 EKRKOKQKLNDRPHEKVPKPR----- 195
Db 152 grqsgerrqlpppapineerrgplpvpvphpgdhgpgsgplslgltvtdiqnppdits 211
QY 196 -----APHDRRREKOKLAQGPPELAEDDANLLHKHIEVANGPASHFETRPQTVD 244
Db 212 sryrglpapgpdpdkkrsgkkiskadigapsg---fkhv-----shvgwbpqngfd 261
QY 245 HMDGSYLSALPFSQ--MSEL-LTRAERVLV-----RPHE--PPPPP 282
Db 262 vnnldpdlrsl-fsragiseaqltdaetsklydfiedqggleavrgemrrqepplpppp 320
QY 283 PMH-----GAGDAKPIPTCISSATGLIENRQSPATGRTPVFVSPTPP----- 325
Db 321 pcrqgggggggggggggggqpl-----rp--pvvsgnkgrrsgplppvpmg 365
QY 326 -PPPPPLPSALSTSSLRASMTSTPPP---VPPPPPPATALQAPVPPPPAPLQIAPGV 381
Db 366 gapppptprg-----pppqrgggpppppppatgrsgp--ppplpgaggp-- 408
QY 382 LHPAPPPIAPPLVQSPVRAAPVCTVPVHPLPQGEVQGLPPPPPPPLPPGIRPSS 441
Db 409 --papp-----pppppppp----- 421
QY 442 PVTVTALAHPPSGLHPTSTAPGPHVPLMPPSPSQVIPASEPKRHPTLTVISDARSVL 501
Db 422 -----pcpgsgap-----pplptpvsgs-----papgggrrgal 452
QY 502 LEATRKGIQLRKVE---EQREQEAKHERIENDVAT---ILSRRIAVEYSDSE-----DD 549
Db 453 ldqirgqilnktptgalensvqppqgqseglvgalhmvmqkrsvihssdegedqtdg 512
QY 550 SEFDEVD 556
Db 513 eeddewd 519

RESULT 15
B43064
ID B43064 standard; Protein; 987 AA.
XX B43064;
AC B43064;
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF2028 polypeptide sequence SEQ ID NO:5656.
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
OS
XX
PN W0200058473-A2.
XX
XX 05-OCT-2000.
PD

Search completed: June 20, 2001, 12:07:35
Job time: 50 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 19:41:43 ; Search time 172.72 Seconds
(without alignments)
8872.313 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 ctctcttgacttgcggat.....aatggttcattttaaagtt 2625

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 678276 seqs, 291890651 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_0401.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
c 1	254	9.7	359	14 Q61275	Human brain Expres
2	177	6.7	242	21 A42628	Human secreted exp
3	110	4.2	165	21 C17657	Human secreted pro
4	23	0.9	359	21 A43416	Rat secreted expre
5	22	0.8	400	21 C10511	Human secreted pro
6	22	0.8	400	21 Z43075	Human 5' EST isola
7	22	0.8	1583	21 Z93303	Sequence encoding
8	22	0.8	2295	21 Z88238	Human cytokine sig
9	21	0.8	1203	21 C36419	Arabidopsis thalia
c 10	21	0.8	3070	21 A26364	Human secreted pro
c 11	21	0.8	4021	21 C77217	Human ORFX ORF2772

c 12	20	0.8	1201	21 A62524	Human secreted pro
c 13	20	0.8	1428	21 A99466	Sakuranetin synth
c 14	20	0.8	3435	17 T35869	Human DNA polymera
c 15	20	0.8	4371	21 A99469	Sakuranetin synth
c 16	20	0.8	5241	21 A99467	Sakuranetin synth
17	19	0.7	300	20 Z13606	Human gene express
18	19	0.7	309	21 C22794	Human secreted pro
19	19	0.7	504	16 T19877	Human gene signatu
20	19	0.7	826	20 V99893	Fragment of aspart
21	19	0.7	927	19 V29596	Humicola insolens
c 22	19	0.7	1210	21 C37485	Arabidopsis thalia
c 23	19	0.7	1257	19 V19378	Humicola insolens
24	19	0.7	1257	19 V13840	Humicola insolens
25	19	0.7	1506	19 V16721	Nucleic acid encod
26	19	0.7	1590	19 V27169	Bacillus thuringie
c 27	19	0.7	1736	21 Z98322	A. thaliana gene i
c 28	19	0.7	1737	20 X99508	Nucleic acid seque
c 29	19	0.7	1839	21 C77337	Human ORFX ORF2892
c 30	19	0.7	2051	19 V66549	Partial thermophil
c 31	19	0.7	2051	21 A38485	DNA encoding therm
c 32	19	0.7	2111	17 T17420	Thermus aquaticus
c 33	19	0.7	2111	19 V66546	DNA encoding & the
c 34	19	0.7	2111	21 A38481	DNA encoding therm
c 35	19	0.7	2181	12 O14967	T. aquaticus ligas
c 36	19	0.7	2352	20 X87656	Thermomyces linugi
37	19	0.7	7720	21 A53800	Genomic DNA encodi
c 38	19	0.7	17341	21 A14872	Maize amylose-exte
c 39	19	0.7	23449	21 Z35393	Minimal motif codi
c 40	18	0.7	24	17 T39968	Multimerisation of
c 41	18	0.7	24	19 V55816	Human secreted pro
c 42	18	0.7	97	21 C12104	Human secreted pro
c 43	18	0.7	98	21 C12043	Tomato genomic DNA
c 44	18	0.7	131	17 T09823	Delta-Taq DNA poly
c 45	18	0.7	160	13 Q24068	

ALIGNMENTS

RESULT 1
Q61275/c
ID Q61275 standard; DNA; 359 BP.
XX Q61275;
XX AC Q61275;
XX DT 16-MAR-1994 (first entry)
XX DE Human brain Expressed Sequence Tag EST01293.
XX KW Gene transcription product; genetic markers; tagging; in vivo;
XX KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX OS Homo sapiens.
XX PN WO9316178-A.
XX PD 19-AUG-1993.
XX PF 12-FEB-1993; 93WO-USO1294.
XX PR 12-FEB-1992; 92US-0837195.
XX

(USSH) US DEPT HEALTH & HUMAN SERVICE.

Adams MD, Moreno RF, Venter CJ;

WPI; 1993-272882/34.

Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes

Example 4; Page 465; 500pp; English.

XX The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST01293 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
XX
XX Sequence 359 BP; 137 A; 57 C; 55 G; 110 T; 0 other;
XX
XX Query Match 9.7%; Score 254; DB 14; Length 359;
XX Best Local Similarity 99.4%; Pred. No. 2e-112;
XX Matches 354; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 2256 aactgaattgaaattcttgaggcagctgtatctactaatgagccttattccatttcctg 2315
XX |||||
XX 356 AACTGAAATTGAAATCTTGAGGCGAGCTGTATCTACTAATGAGCCTTATTCATTCTCTG 297
XX |||||
XX 2316 atgttttaaaagaagaacactgccttgattatatacgaatacaactcagaagaatatttag 2375
XX |||||
XX 296 ATGTTTAAAGAGGAGAAACACTGCCCTTGATTATACGAATACACTCAGAAAGTACATTAG 237
XX |||||
XX 2376 ctgttagtgaattctctaaagaagctgtgaatttttcattattttattatt 2435
XX |||||
XX 236 CTGTGAGTGTGAATCTCTTAAGAGAAATGCTTGAAATTTTTCATTATGTTTATTGGT 177
XX |||||
XX 2436 ttatatacttgccttatttgaatttttagcagtgatcccttcccacttatatttgg 2495
XX |||||
XX 176 TTTATATACTTGCCTTAATTGAATCTTGAAGATATCCCTCCACATATATATGTTGT 117
XX |||||
XX 2496 gatagtttgccttgcctatagagtgtaaaacttttccatgtaaatctctgaactta 2555
XX |||||
XX 116 GATATGATTTTCTGCTTATAGGAGTGTAAACATTTTCCATGTGAAATCTCTGACATTA 57
XX |||||
XX 2556 aacatacatgaactacataactgttgaagaatacaactcgtgatttaataatggt 2611
XX |||||
XX 56 AACATACATGTACATCTTACTACTTGTAGGAAATACAGCTGCTGATTTAATATGCT 1

XX 15-OCT-1999; 99WO-US24205.
XX PF
XX 15-OCT-1998; 98US-0104435.
XX PR
XX (GEMV) GENETICS INST INC.
XX PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M;
XX PT WPI; 2000-317937/27.
XX DR
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX PT expressed sequence tags (seSTs), useful for treating various disorders
XX PT such as autoimmune, infectious, and central nervous system disorders.
XX PT
XX Claim 1; Page 450; 618pp; English.
XX PS
XX A41261 to A43419 represent specifically claimed secreted expressed
XX CC sequence tags (seSTs), isolated from human, mouse, xenopus and rat
XX CC tissue sources. The seSTs can have a range of activities depending on
XX CC the tissues they were isolated from. The activities include: chemotactic;
XX CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
XX CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
XX CC antifungal; antiviral; antidiabetic; antidiabetic; vulnery; antitumor;
XX CC osteoprotective; neuroprotective; nontropic; antiparkinsonian; antipsoriatic;
XX CC cerebroprotective; anticonvulsant; and antidepressant. The seSTs can be
XX CC used for gene therapy and in vaccines. The seSTs are useful as probes for
XX CC the identification and isolation of full-length cDNAs and genomic DNA
XX CC molecules which correspond to the seSTs. Proteins encoded by the seSTs
XX CC are useful in assays for determining biological activity and raising
XX CC antibodies. They may be useful for treatment of autoimmune disorders
XX CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX CC osteoporosis, osteoarthritis, central nervous system disorders
XX CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX CC disease), tumours, bacterial, fungal or viral infections, depression
XX CC and psoriasis. A4320 to A43425 represent linker variants which are
XX CC given in the exemplification of the present invention.
XX
XX Sequence 242 BP; 83 A; 52 C; 65 G; 42 T; 0 other;
XX
XX Query Match 6.7%; Score 177; DB 21; Length 242;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-75;
XX Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 784 agaaaatcttagatcgtctctcatgaaccagaaaagtgccaaagagcaccctcatgacagcc 843
XX |||||
XX 63 agaaaatcttagatcgtctctcatgaaccagaaaagtgccaaagagcaccctcatgacagcc 122
XX |||||
XX 844 ggcagaatggcagaagctggcccaaggtccagagctggcgaagatgatgtaattctct 903
XX |||||
XX 123 ggcagaatggcagaagctggcccaaggtccagagctggcgaagatgatgtaattctct 182
XX |||||
XX 904 tacataagcatattgaagtgtctaataagccagcctctcattttgaacaagacctc 960
XX |||||
XX 183 tacataagcatattgaagtgtctaataagccagcctctcattttgaacaagacctc 239
XX |||||
XX
XX RESULT 3
XX C17657
XX ID C17657 standard; cDNA; 165 BP.
XX XX
XX C17657;
XX AC
XX 06-OCT-2000 (first entry)
XX DT
XX Human secreted protein 5' EST, SEQ ID NO: 21732.
XX DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX KW

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 XX 06-SEP-2000.
 XX
 XX 21-FEB-2000; 2000EP-0200610.
 PF
 XX 26-FEB-1999; 99US-0122487.
 PR
 XX (GEST) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 DR
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PT
 XX Claim 1; SEQ ID 14586; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 XX Sequence 400 BP; 106 A; 111 C; 110 G; 70 T; 3 other;
 SQ
 Query Match 0.8%; Score 22; DB 21; Length 400;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1075 cactccacccaccacgaatgca 1096
 Db 254 cactccacccaccacgaatgca 275
 RESULT 6
 Z43075
 ID Z43075 standard; cDNA; 400 BP.
 AC Z43075;
 XX
 XX 01-FEB-2000 (first entry)
 DT
 XX Human 5' EST isolated from a cDNA library SEQ ID NO:1622.
 DE
 XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification; ss.
 XX
 OS Homo sapiens.
 XX
 XX W09953051-A2.
 PN
 XX 21-OCT-1999.
 PD

XX 09-APR-1999; 99WO-IB00712.
 PF
 XX 09-APR-1998; 98US-0057719.
 PR
 XX 28-APR-1998; 98US-0069047.
 XX
 XX (GEST) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-038446/03.
 DR
 XX Novel secreted protein 5' expressed sequence tag sequences used in
 XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
 PT
 XX Claim 1; Page 837; 837pp; English.
 PS
 XX Z42265 to Z43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. Y64651 to Y65438
 CC represent the EST-related proteins corresponding to Z42265 to Z43052.
 CC The 5' ESTs can be used for producing secreted human gene products.
 CC They can be used to identify and isolate 5' untranslated regions (UTRs)
 CC and upstream regulatory regions which control the location, development
 CC stage, rate, and quantity of protein synthesis, as well as stability of
 CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
 CC obtain full length cDNA clones. The ESTs can also be used in forensic
 CC procedures to identify individuals, or in diagnostic procedures to
 CC identify individuals having genetic diseases resulting from abnormal
 CC gene expression. The products may also be used in gene therapy protocols.
 CC The nucleic acids encoding signal peptides can be used for directing
 CC extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.
 CC The proteins encoded by the EST sequences may be useful in treating a
 CC variety of human conditions. Secreted proteins have therapeutic value,
 CC and the identification of new secreted proteins is valuable. Z42269 to
 CC Z42264 and Y64644 to Y64650 represent sequences used in the
 CC exemplification of the present invention.
 XX
 XX Sequence 400 BP; 106 A; 111 C; 110 G; 70 T; 3 other;
 SQ
 Query Match 0.8%; Score 22; DB 21; Length 400;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1075 cactccacccaccacgaatgca 1096
 Db 254 cactccacccaccacgaatgca 275
 RESULT 7
 Z93303
 ID Z93303 standard; DNA; 1583 BP.
 AC Z93303;
 XX
 XX 04-JUL-2000 (first entry)
 DT
 XX Sequence encoding human homer interacting protein I30.
 DE
 XX Homer; calcium; receptor; immediate early gene; IEG;
 KW identification; treatment; glutamate receptor;
 KW inositol triphosphate; epilepsy; glutamate toxicity;
 KW memory disorder; learning disorder; stroke; schizophrenia;
 KW Alzheimer's disease; tissue degeneration; brain development;
 KW cardiac disorder; muscular disorder; vascular disorder;
 KW neurological disorder; psychiatric disorder; renal disorder;
 KW uterine disorder; bronchial disorder; ageing; human; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 160..1120
 FT

XX Arabidopsis thaliana.
OS EP1033405-A2.
PN 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-012180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135622.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.8%; Score 21; DB 21; Length 1203;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1211 tccccaccctctccaccacc 1231
|||||
DB 493 tccccaccctctccaccacc 513

RESULT 10
A26364/c
ID A26364 standard; cDNA; 3070 BP.
XX AC A26364;
XX AC A26364;
DT 29-JUN-2000 (first entry)
XX XX Human secreted protein gene 19 SEQ ID NO:29.
DE DE Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; neutropenic; neuroprotective; antiasthmatic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.
XX OS Homo sapiens.
XX OS Homo sapiens.
PN W0200006698-A1.
XX XX

PD 10-FEB-2000.
XX 29-JUL-1999; 99WO-US17130.
XX 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Komatsoulis GA, Rösen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX WPI; 2000-195282/17.
DR P-PSDB; Y91469.
XX
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders -
XX
XX Claim 1; Page 385; 634pp; English.
XX
CC The polynucleotide sequences given in A26346 to A26458 encode the human
CC secreted proteins given in Y91451 to Y91691. The human secreted proteins
CC can have activities based on the tissues and cells they are expressed in.
CC Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
CC antiinflammatory; neutropenic; neuroprotective; antiasthmatic; osteopathic;
CC antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic;
CC and cardiant. The polynucleotides and their corresponding secreted
CC proteins are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the proteins in a sample or
CC by determining the presence of mutations in the polynucleotides. Specific
CC uses are described for each of the polynucleotides, based on which
CC tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and foetal
CC deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The proteins
CC or polynucleotides can also be used as food additives or preservatives.
CC The proteins are also useful for identifying their binding partners.
CC A26337 to A26345 and Y91450 are sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 3070 BP; 781 A; 725 C; 619 G; 945 T; 0 other;

Query Match 0.8%; Score 21; DB 21; Length 3070;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2402 aatgcttgattttttcatta 2422
|||||
DB 1603 AATGCTTGAATTTTTCATTA 1583

RESULT 11
C77217/c
ID C77217 standard; cDNA; 4021 BP.
XX AC C77217;
XX AC C77217;
DT 08-FEB-2001 (first entry)
XX XX Human ORF2772 polynucleotide sequence SEQ ID NO:5543.
DE DE

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparisonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antithyroid; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disease; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 XX
 XX Homo sapiens.
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 PI Shimkets RA, Leach M;
 XX
 XX WPI: 2000-602362/57.
 DR P-PSDB; B43008.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4726-4728; 5507pp; English.
 XX
 CC C74446 to C77606 encode the proteins given in B40237 to B43397, which
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparisonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antithyroid; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 4021 BP; 821 A; 1255 C; 1220 G; 723 T; 2 other;

Query Match 0.8%; Score 21; DB 21; Length 4021;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1505 ggggctgctccaccaccac 1525
 Db 411 GGGGCTGCTCCACCCACC 391

RESULT 12
 A62524/c
 ID A62524 standard; cDNA; 1201 BP.
 XX
 AC A62524;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Human secreted protein cDNA clone 3903091.
 XX
 KW Human; secreted protein; SECX; cancer; cytostatic; vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 174..1193
 FT /tag= a
 FT /product= "SECX polypeptide"
 FT /transl_except= (pos:999..1001,aa:Xaa)
 FT /note= "Xaa= unknown"
 FT misc_feature 174..1190
 FT /tag= b
 FT /note= "given in the specification as SEQ ID NO: 119
 FT and specifically claimed in claim 4"
 FT sig_peptide 174..224
 FT /tag= c
 FT mat_peptide 225..1190
 FT /tag= d
 XX
 PN WO200037634-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 17-DEC-1999; 99WO-US29854.
 XX
 PR 18-DEC-1998; 98US-0112837.
 PR 21-DEC-1998; 98US-0113485.
 PR 16-DEC-1999; 99US-0113485.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 PI Shimkets RA;
 XX
 DR WPI: 2000-442664/38.
 DR P-PSDB; B14313.
 XX
 PT New SECX polypeptide useful for diagnosing, treating and delaying
 PT pathological conditions, especially cancer -
 XX
 PS Claim 5; Page 118-120; 155pp; English.
 XX
 CC The present sequence is one of 23 nucleic acids, collectively referred to
 CC as SECX nucleic acids, which encode novel secreted human proteins.
 CC Various SECX nucleic acids were cloned into insect cells and human
 CC embryonic kidney 293 cells and quantitative expression of the clones
 CC was analysed. By detecting the presence of a SECX nucleic acid
 CC or polypeptide in a sample it is possible to diagnose a pathological
 CC condition, especially cancer, associated with aberrant SECX expression or
 CC activity. Antibodies, and primers and probes are used, respectively, to
 CC detect SECX polypeptides and SECX nucleic acids in this method.
 CC SECX nucleic acids, polypeptides and antibodies are also useful for
 CC treating, preventing or delaying conditions associated with aberrant SECX
 CC expression.
 XX
 SQ Sequence 1201 BP; 438 A; 201 C; 219 G; 337 T; 6 other;

Query Match 0.8%; Score 20; DB 21; Length 1201;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2408 tgaatttttccattattgtt 2427
|||||
Db 186 TGAATTTTTCATTATGTT 167

RESULT 13

A99466/c
ID A99466 standard; DNA; 1428 BP.

XX AC A99466;

DT 26-JAN-2001 (first entry)

XX DE Sakuranetin synthase gene.

XX KW Sakuranetin synthetase; naringenin 7-O-methyltransferase; NOMT;

XX KW antibacterial; transgenic plant; ds.

XX OS Oryza sativa.

XX XX WO200052174-A1.

XX PN 08-SEP-2000.

XX PF 03-MAR-2000; 2000WO-JP01306.

XX PR 04-MAR-1999; 99JP-0057748.

XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX PI Kodama O;

XX DR WPI: 2000-572186/53.

XX DR P-PSDB; B26843.

XX PT Sakuranetin synthetase gene encoding protein with naringenin
PT 7-O-methyltransferase activity in rice cells and promoter activity,
PT used to derive sakuranetin from naringenin easily in plants to impart
PT antibacterial characteristics -

XX PS Claim 9; Page 36-37; 57pp; Japanese.

XX CC This invention relates to a sakuranetin synthetase gene encoding a
CC protein with naringenin 7-O-methyltransferase (NOMT) activity and
CC promoter activity. The invention includes a recombinant vector including
CC the sakuranetin synthetase DNA and host cells transformed with the
CC vector. The sakuranetin synthetase protein has antibacterial activity and
CC is used to produce transgenic plants with antibacterial characteristics.
CC The present sequence represents genomic DNA encoding the enzyme of the
CC invention.

XX SQ Sequence 1428 BP; 496 A; 292 C; 226 G; 414 T; 0 other;

Query Match 0.8%; Score 20; DB 21; Length 1428;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2007 tgttttgcctttcttccctt 2026
|||||

Db 1227 TGTTTTGCCTTTCTTCCCTT 1208

RESULT 14

T35869/c

ID T35869 standard; cDNA; 3435 BP.

XX AC T35869;

XX DT 21-OCT-1996 (first entry)

XX DE Human DNA polymerase delta cDNA.

XX KW DNA polymerase delta; pol delta; colorectal tumour; cancer;
XX gene therapy; diagnosis; replication error; RER+; DNA repair; ds.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX CDS 43..3366
XX FT /*tag= a

XX PN WO9621026-A1.

XX PD 11-JUL-1996.

XX PF 02-JAN-1996; 96WO-US000005.

XX PR 30-DEC-1994; 94US-0366577.

XX PA (UYJO) UNIV JOHNS HOPKINS.

XX PI Kinzler KW, Vogelstein B;

XX DR WPI: 1996-334001/33.

XX DR P-PSDB; W03132.

XX PT DNA polymerase delta mutant and corresp. coding sequence -
PT associated with replication error phenotype in colorectal tumours,
PT useful for diagnosis and therapy

XX PS Claim 1; Page 13-17; 32pp; English.

XX CC A cDNA clone (T35869) corresponding to the human pol delta gene
CC codes for DNA polymerase delta (W03132), an enzyme which has 3'-5'
CC exonuclease activity. Mutations of the gene lead to a replication
CC error phenotype (RER+) in some colorectal cancers. Variants at
CC codon 502 (proximal to the ExoIII domain) and codon 506 (within
CC ExoIII) of the gene were separately found in 2 colorectal cancer
CC cell lines (see also W03133-34). Persons predisposed to hereditary
CC non-polyposis colorectal cancer, owing to a mutation in a pol delta
CC allele, can be treated by administration of the gene. Patients
CC with RER+ tumours can be treated to prevent accumulation of somatic
CC mutations leading to resistance to cytostatic agents. The gene
CC can also be used to diagnose a predisposition to colorectal cancer.

XX SQ Sequence 3435 BP; 621 A; 1108 C; 1115 G; 591 T; 0 other;

Query Match 0.8%; Score 20; DB 17; Length 3435;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1651 tgctcccatctcctccatca 1670
|||||

Db 180 TGCTCCCATCTCTCCCATCA 161

RESULT 15

A99469/c

ID A99469 standard; DNA; 4371 BP.

XX AC A99469;

XX DT 26-JAN-2001 (first entry)

XX DE Sakuranetin synthase genomic DNA sequence.

XX KW Sakuranetin synthetase; naringenin 7-O-methyltransferase; NOMT;
XX antibacterial; transgenic plant; ds.

XX OS Oryza sativa.

XX PN WO200052174-A1.

XX XX

PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-JP01306.
XX
PR 04-MAR-1999; 99JP-0057748.
XX
PA (NTSC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Kodama O;
XX
DR WPI; 2000-572186/53.
XX
PT Sakuranetin synthetase gene encoding protein with naringenin
7-O-methyltransferase activity in rice cells and promoter activity,
PT used to derive sakuranetin from naringenin easily in plants to impart
antibacterial characteristics -
XX
PS Example 3; Page 33-36; 57pp; Japanese.
XX
CC This invention relates to a sakuranetin synthetase gene encoding a
protein with naringenin 7-O-methyltransferase (NOMT) activity and
promoter activity. The invention includes a recombinant vector including
the sakuranetin synthetase DNA and host cells transformed with the
vector. The sakuranetin synthetase protein has antibacterial activity and
is used to produce transgenic plants with antibacterial characteristics.
CC The present sequence represents the sakuranetin synthetase gene of the
invention.
XX
SQ Sequence 4371 BP; 1281 A; 1008 C; 949 G; 1133 T; 0 Other;

Query Match 0.8%; Score 20; DB 21; Length 4371;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2007 tggtttgctttctttcctt 2026
|||||
Db 1227 TGTTTTGCTTTCTTTCTT 1208

Search completed: June 20, 2001, 23:25:27
Job time: 13424 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 19:27:34 ; Search time 89.51 Seconds
(without alignments)
5433.006 Million cell updates/sec

Title: US-09-425-501-1

Perfect score: 2625

Sequence: 1 ctctcttgcacttcgcat.....aatggttcattttaaaagt 2625

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 317530 seqs, 92630169 residues

Word size : 0

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:**

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:**
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:**
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:**
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:**
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:**
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	0.8	2295	3	US-09-189-035-3
2	22	0.8	2295	4	US-09-382-086-3
c 3	20	0.8	3435	1	US-08-366-577-1
c 4	20	0.8	3435	5	PCT-US96-00005-1
5	19	0.7	807	3	US-08-718-905-1
6	19	0.7	1257	4	US-09-230-222-2
c 7	19	0.7	2051	1	US-08-343-785-7
c 8	19	0.7	2051	2	US-08-462-221-7
c 9	19	0.7	2051	3	US-08-946-458-7
c 10	19	0.7	2111	1	US-08-343-785-1
c 11	19	0.7	2111	2	US-08-462-221-1
c 12	19	0.7	2111	3	US-08-946-458-1
c 13	18	0.7	24	2	US-08-529-1908-8
c 14	18	0.7	131	1	US-08-248-474-102
c 15	18	0.7	131	3	US-08-756-849-102
c 16	18	0.7	160	1	US-08-274-205-2
c 17	18	0.7	962	1	US-08-073-3846-11
c 18	18	0.7	962	1	US-08-254-359A-11
c 19	18	0.7	962	1	US-08-483-043-11
c 20	18	0.7	962	1	US-08-481-238-11
c 21	18	0.7	962	2	US-08-471-066B-11
c 22	18	0.7	962	2	US-08-484-956-11
c 23	18	0.7	962	2	US-08-757-653-11
c 24	18	0.7	962	2	US-08-599-491-11
c 25	18	0.7	962	2	US-08-756-386-11
c 26	18	0.7	962	2	US-08-823-516-11
c 27	18	0.7	962	3	US-08-682-853A-11

c 28	18	0.7	962	3	US-08-759-038-11	Sequence 11, Appl
c 29	18	0.7	962	3	US-08-758-314-11	Sequence 11, Appl
c 30	18	0.7	969	1	US-08-254-359A-30	Sequence 30, Appl
c 31	18	0.7	969	2	US-08-471-066B-30	Sequence 30, Appl
c 32	18	0.7	969	2	US-08-484-956-30	Sequence 30, Appl
c 33	18	0.7	969	2	US-08-757-653-30	Sequence 30, Appl
c 34	18	0.7	969	2	US-08-599-491-30	Sequence 30, Appl
c 35	18	0.7	969	2	US-08-756-386-30	Sequence 30, Appl
c 36	18	0.7	969	2	US-08-823-516-25	Sequence 25, Appl
c 37	18	0.7	969	3	US-08-682-853A-30	Sequence 30, Appl
c 38	18	0.7	969	3	US-08-759-038-30	Sequence 30, Appl
c 39	18	0.7	969	3	US-08-758-314-30	Sequence 30, Appl
c 40	18	0.7	1600	1	US-08-073-384C-12	Sequence 12, Appl
c 41	18	0.7	1600	1	US-08-254-359A-12	Sequence 12, Appl
c 42	18	0.7	1600	1	US-08-483-043-12	Sequence 12, Appl
c 43	18	0.7	1600	1	US-08-481-238-12	Sequence 12, Appl
c 44	18	0.7	1600	2	US-08-471-066B-12	Sequence 12, Appl
c 45	18	0.7	1600	2	US-08-484-956-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-189-035-3
; Sequence 3, Application US/09189035
; Patent No. 6020165
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Gueghen, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
; FILE REFERENCE: PF-0638 US
; CURRENT APPLICATION NUMBER: US/09/189,035
; CURRENT FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2280326
US-09-189-035-3

Query Match 0.8%; Score 22; DB 3; Length 2295;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1505 ggggctgcctccacccaccca 1526
|||||
DB 1367 ggggctgcctccacccaccca 1388
RESULT 2
US-09-382-086-3
; Sequence 3, Application US/09382086
; Patent No. 6201106
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Gueghen, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
; FILE REFERENCE: PF-0638 US
; CURRENT APPLICATION NUMBER: US/09/382,086
; CURRENT FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 09/189,035
; EARLIER FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program

SEQ ID NO 3
LENGTH: 2295
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 2280326
US-09-382-086-3

Query Match 0.8%; Score 22; DB 4; Length 2295;
Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Indels 0; Gaps 0;

Qy 1505 ggggctgcctcccccaccca 1526
Db 1367 ggggctgcctcccccaccca 1388

RESULT 3
US-08-366-577-1/c
Sequence 1, Application US/08366577
Patent No. 5728523
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth W.
TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL
TITLE OF INVENTION: TUMORS WITH REPLICATION ERRORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366, 577
FILING DATE: 12-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.48554
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43...3364
US-08-366-577-1

Query Match 0.8%; Score 20; DB 1; Length 3435;
Best Local Similarity 100.0%; Pred. No. 3.2; Mismatches 0; Indels 0; Gaps 0;

Qy 1651 tgctccatctctccatca 1670
Db 180 TGCCTCCATCTCTCCATCA 161

RESULT 4
PCT-US96-00005-1/c
Sequence 1, Application PC/TUS9600005
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth W.
TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL
TITLE OF INVENTION: TUMORS WITH REPLICATION ERRORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, Ltd.
STREET: 1001 G Street N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00005
FILING DATE: 2-JAN-96
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.53505
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 43...3364
PCT-US96-00005-1

Query Match 0.8%; Score 20; DB 5; Length 3435;
Best Local Similarity 100.0%; Pred. No. 3.2; Mismatches 20; Conservative 0; Indels 0; Gaps 0;

Qy 1651 tgctccatctctccatca 1670
Db 180 TGCCTCCATCTCTCCATCA 161

RESULT 5
US-08-718-905-1
Sequence 1, Application US/08718905
Patent No. 6063756
GENERAL INFORMATION:
APPLICANT: Donovan, William P.
APPLICANT: Donovan, Judith C.
APPLICANT: Slaney, Annette C.
TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET33 AND CRYET34

;; TITLE OF INVENTION: COMPOSITIONS AND USES THEREFOR
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Arnold, White & Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: Texas
;; COUNTRY: United States of America
;; ZIP: 77210
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/718,905
;; FILING DATE: Concurrently Herewith
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kitchell, Barbara S.
;; REGISTRATION NUMBER: 33,928
;; REFERENCE/DOCKET NUMBER: MOBT:003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 418-3000
;; TELEFAX: (512) 474-7577
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 807 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-718-905-1

Query Match 0.7%; Score 19; DB 3; Length 807;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 2598 atttaataatgggttcatt 2616
Db 743 ATTTAATAATGGTTCATT 761

RESULT 6
US-09-230-222-2
; Sequence 2, Application US/09230222A
; Patent No. 6159720
; GENERAL INFORMATION:
; APPLICANT: MURASHIMA, KOUICHIROU
; APPLICANT: MORIYA, TATSUKI
; APPLICANT: HAMAYA, TORU
; APPLICANT: KOGA, JINICHIRO
; APPLICANT: SUMIDA, NAOMI
; APPLICANT: AOYAGI, KAORU
; APPLICANT: MURAKAMI, TAKESHI
; APPLICANT: KONO, TOSHIKI
; TITLE OF INVENTION: ENZYME ENDOGLUCANASE AND CELLULASE PREPARATIONS
; FILE REFERENCE: 99-0055*/LC(WMC)/144
; CURRENT APPLICATION NUMBER: US/09/230,222A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Humicola insolens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (118)..(180)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)..(180)

;; FEATURE:
;; NAME/KEY: mat_peptide
;; LOCATION: (181)..(1088)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (181)..(452)
;; FEATURE:
;; NAME/KEY: Intron
;; LOCATION: (453)..(508)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (509)..(1088)
;; US-09-230-222-2

Query Match 0.7%; Score 19; DB 4; Length 1257;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1073 accacotccacotccacca 1091
Db 921 accacotccacotccacca 939

RESULT 7
US-08-343-785-7/c
; Sequence 7, Application US/08343785
; Patent No. 5494810
; GENERAL INFORMATION:
; APPLICANT: Francis Barany et al
; TITLE OF INVENTION: Thermostable Ligase Mediated DNA
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,785
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/971,095
; FILING DATE: No. 5494810ember 2nd 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: CRF D-933/CIT-2047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-343-785-7

Query Match 0.7%; Score 19; DB 1; Length 2051;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1210 ctccccacacctctccacc 1228
|||||
Db 1665 CTCCTCCACCTCTCCACC 1647

RESULT 8

US-08-462-221-7/c

; Sequence 7, Application US/08462221

; Patent No. 5830711

; GENERAL INFORMATION:

; APPLICANT: Barany, Francis

; APPLICANT: Zebala, John

; APPLICANT: Nickerson, Deborah

; APPLICANT: Kaiser, Jr., Robert J.

; APPLICANT: Hood, Leroy

; TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA

; TITLE OF INVENTION: AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC DISEASES

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/462,221

; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/343,785

; FILING DATE: 22-NOV-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,095

; FILING DATE: 02-NOV-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/518,447

; FILING DATE: 03-MAY-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/267

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 716-263-1304

; TELEFAX: 716-263-1600

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2051 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

US-08-462-221-7

Query Match

Best Local Similarity 0.7%; Score 19; DB 2; Length 2051;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1210 ctccccacacctctccacc 1228
|||||
Db 1665 CTCCTCCACCTCTCCACC 1647

RESULT 9

US-08-946-458-7/c

; Sequence 7, Application US/08946458

; Patent No. 6054564

; GENERAL INFORMATION:
; APPLICANT: Barany, Francis
; APPLICANT: Zebala, John
; APPLICANT: Nickerson, Deborah
; APPLICANT: Kaiser, Jr., Robert J.
; APPLICANT: Hood, Leroy
; TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA
; TITLE OF INVENTION: AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC DISEASES
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon Peabody LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,221
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: US 08/343,785
; FILING DATE: 22-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,095
; FILING DATE: 02-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,447
; FILING DATE: 03-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/267
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-946-458-7

Query Match 0.7%; Score 19; DB 3; Length 2051;

Best Local Similarity 100.0%; Pred. No. 9.7;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1210 ctccccacacctctccacc 1228
|||||
Db 1665 CTCCTCCACCTCTCCACC 1647

RESULT 10

US-08-343-785-1/c

; Sequence 1, Application US/08343785

; Patent No. 5494810

; GENERAL INFORMATION:

; APPLICANT: Francis Barany et al

; TITLE OF INVENTION: Thermostable Ligase Mediated DNA

; TITLE OF INVENTION: Amplification System For The

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Yahwak & Associates

STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,785
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/971,095
FILING DATE: NO. 5494810ember 2nd 1992
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-933/CIT-2047
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-343-785-1

Query Match 0.7%; Score 19; DB 1; Length 2111;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 ctccccacacctctccacc 1228
|||||
DB 1725 CTCCTCCACCTCTCTCCACC 1707

RESULT 11

US-08-462-221-1/c
Sequence 1, Application US/08462221
Patent No. 5830711
GENERAL INFORMATION:
APPLICANT: Barany, Francis
APPLICANT: Zebala, John
APPLICANT: Nickerson, Deborah
APPLICANT: Kaiser, Jr., Robert J.
APPLICANT: Hood, Leroy
TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA
TITLE OF INVENTION: AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC DISEASES
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,221
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/343,785
FILING DATE: 22-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,095
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,447
FILING DATE: 03-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/267
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-462-221-1

Query Match 0.7%; Score 19; DB 2; Length 2111;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 ctccccacacctctccacc 1228
|||||
DB 1725 CTCCTCCACCTCTCTCCACC 1707

RESULT 12

US-08-946-458-1/c
Sequence 1, Application US/08946458
Patent No. 6054564
GENERAL INFORMATION:
APPLICANT: Barany, Francis
APPLICANT: Zebala, John
APPLICANT: Nickerson, Deborah
APPLICANT: Kaiser, Jr., Robert J.
APPLICANT: Hood, Leroy
TITLE OF INVENTION: A THERMOSTABLE LIGASE MEDIATED DNA
TITLE OF INVENTION: AMPLIFICATION SYSTEM FOR THE DETECTION OF GENETIC DISEASES
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon Peabody LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,458
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/462,221
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/343,785
FILING DATE: 22-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,095
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,447

;; FILING DATE: 03-MAY-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Goldman, Michael L.
;; REGISTRATION NUMBER: 30,727
;; REFERENCE/DOCKET NUMBER: 19603/267
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 716-263-1304
;; TELEFAX: 716-263-1600
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2111 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-946-458-1

Query Match 0.7%; Score 19; DB 3; Length 2111;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1210 ctcccccaactctccacc 1328
Db 1725 CTCCTCCACCTCTCCACC 1707

RESULT 13
US-08-529-190B-8/c
; Sequence 8, Application US/08529190B
; Patent No. 5833991
; GENERAL INFORMATION:
; APPLICANT: Masucci, Maria G.
; TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES
; TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,190B
; FILING DATE: 15-SEP-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE9501324-9
; FILING DATE: 10-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US08/522,595
; FILING DATE: 01-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen A
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3255/53015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-345-9100
; TELEFAX: 617-345-9111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
US-08-529-190B-8

Query Match 0.7%; Score 18; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1351 cacctccagctccctcttc 1368
Db 19 CACCTCCAGCTCCCTCTTC 2

RESULT 14
US-08-248-474-102/c
; Sequence 102, Application US/08248474
; Patent No. 5612471
; GENERAL INFORMATION:
; APPLICANT: MCK, BIRD, David
; APPLICANT: WILSON, Mark A.
; TITLE OF INVENTION: NEMATODE-INDUCE GENES IN TOMATO
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,474
; FILING DATE: 25-MAY-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 2307E-535
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Lycopersicon esculentum cv 'Rutgers Large
; ORGANISM: Red'
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..131
; OTHER INFORMATION: //standard_name= "DB# 275"
US-08-248-474-102

Query Match 0.7%; Score 18; DB 1; Length 131;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1305 ccactccactccagcc 1322
Db 105 CCACCTCCACCTCCAGCC 88

RESULT 15
US-08-756-849-102/c
; Sequence 102, Application US/08756849
; Patent No. 6093810


```

; GENERAL INFORMATION:
; APPLICANT: Bird, David Mck.
; APPLICANT: Wilson, Mark A.
; TITLE OF INVENTION: Nematode-Induced Genes in Tomato
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,849
; FILING DATE: 26-NOV-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,474
; FILING DATE: 25-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-053510US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Lycopersicon esculentum cv 'Rutgers Large Red'
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..131
; OTHER INFORMATION: /standard_name= "DB# 275"
; US-08-756-849-102

```

```

Query Match      0.78; Score 18; DB 3; Length 131;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1305 ccacctccacctccagcc 1322
    |
Db 105 CCACCTCCACCTCCAGCC 88

```

Search completed: June 20, 2001, 23:22:20
Job time: 14086 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 17:22:49 ; Search time 1942.94 Seconds
(without alignments)
12771.234 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 cttctctgacttgaggat.....aatgggttcattttaaaagtt 2625

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est33:*
- 26: gb_est34:*
- 27: gb_est35:*
- 28: gb_est36:*
- 29: gb_est37:*
- 30: gb_est38:*
- 31: gb_est39:*
- 32: gb_est40:*
- 33: em_estba:*
- 34: em_estfun:*
- 35: em_esthum1:*
- 36: em_esthum2:*
- 37: em_esthum3:*
- 38: em_esthum4:*
- 39: em_esthum5:*
- 40: em_esthum6:*
- 41: em_esthum7:*
- 42: em_esthum8:*
- 43: em_esthum9:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estin5:*
- 68: em_estom1:*
- 69: em_estom2:*
- 70: em_estov1:*
- 71: em_estov2:*
- 72: em_estp11:*
- 73: em_estp12:*
- 74: em_estp13:*
- 75: em_estp14:*
- 76: em_estp15:*
- 77: em_estp16:*
- 78: em_estp17:*
- 79: em_estp18:*
- 80: em_estp19:*
- 81: em_estp110:*
- 82: em_estro1:*
- 83: em_estro2:*
- 84: em_estro3:*
- 85: em_estro4:*
- 86: em_estro5:*
- 87: em_estro6:*
- 88: em_estro7:*
- 89: em_estro8:*
- 90: em_estro9:*
- 91: em_estro10:*
- 92: em_estro11:*
- 93: em_estro12:*
- 94: em_estro13:*
- 95: em_estro14:*
- 96: em_estro15:*
- 97: em_estro16:*
- 98: em_estro17:*
- 99: em_estro18:*
- 100: em_estro19:*
- 101: em_estro20:*
- 102: gb_est25:*
- 103: gb_est26:*
- 104: gb_est27:*
- 105: gb_est28:*
- 106: gb_est29:*
- 107: gb_est30:*
- 108: gb_est31:*
- 109: gb_est32:*
- 110: gb_est41:*
- 111: gb_est42:*
- 112: gb_est43:*
- 113: gb_est44:*
- 114: gb_est45:*
- 115: gb_est46:*
- 116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	585	22.3	851	105	AL526626	AL526626
2	561	21.4	652	175	BG288785	BG288785
3	543	20.7	636	122	AW963370	AW963370
C 4	541	20.6	718	115	AW379048	AW379048
C 5	496	18.2	497	111	AW071606	AW071606
C 6	479	18.2	540	103	AI928895	AI928895
7	467	17.8	700	108	AU142666	AU142666
8	461	17.6	821	154	BG492540	BG492540
C 9	460	17.5	462	23	AI650710	AI650710
C 10	432	16.5	555	164	BE148072	BE148072
C 11	430	16.4	430	119	AW664122	AW664122
C 12	398	15.2	565	108	AU159926	AU159926
C 13	398	15.2	727	139	BE731111	BE731111
C 14	391	14.9	443	187	N95702	N95702
C 15	389	14.8	481	17	AI223407	AI223407
C 16	387	14.7	394	24	AI750422	AI750422
C 17	383	14.6	388	119	AW673234	AW673234
C 18	381	14.5	988	173	BG111973	BG111973
C 19	379	14.4	692	105	AL079676	AL079676
C 20	349	13.3	367	153	BG390996	BG390996
C 21	349	13.3	367	153	BG391003	BG391003
C 22	349	13.3	367	153	BG391089	BG391089
C 23	347	13.2	371	116	AW451579	AW451579
C 24	335	12.8	395	15	AI061122	AI061122
C 25	335	12.8	451	13	AA885055	AA885055
C 26	329	12.5	600	103	AI929146	AI929146
C 27	327	12.5	878	153	BG426480	BG426480
C 28	320	12.2	354	187	R56489	R56489
C 29	315	12.0	1142	174	BG168604	BG168604
C 30	314	12.0	434	157	HJ8215	HJ8215
C 31	311	11.8	311	118	AW590436	AW590436
C 32	304	11.6	406	12	AA846712	AA846712
C 33	301	11.5	650	122	AW956469	AW956469
C 34	289	11.0	397	18	AI306688	AI306688
C 35	279	10.6	321	157	D57242	D57242
C 36	274	10.4	322	19	AI341434	AI341434
C 37	272	10.4	409	103	AI910678	AI910678
C 38	267	10.2	357	157	H06129	H06129
C 39	262	10.0	778	140	BE787088	BE787088
C 40	255	9.7	430	162	BE019589	BE019589
C 41	254	9.7	330	157	F12990	F12990
C 42	254	9.7	359	159	M79145	M79145
C 43	253	9.6	390	8	AA515859	AA515859
C 44	252	9.6	394	159	N63410	N63410
C 45	251	9.6	492	136	BE463994	BE463994

ALIGNMENTS

RESULT 1	
AL526626	
LOCUS	851 bp mRNA EST 13-FEB-2001
DEFINITION	AL526626 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC020YO18 5 prime, mRNA sequence.
ACCESSION	AL526626
VERSION	AL526626.1 GI:12790119
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope

Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.	
Location/Qualifiers	
FEATURES	source
1..851	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="CS0DC020YO18"
	/clone_lib="LTI_NFL003_NBC3"
	/sex="male"
	/tissue_type="neuroblastoma cells"
	/lab_host="DH10B"
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT	263 a 132 c 153 g 299 t 4 others
ORIGIN	
Query Match	22.3%; Score 585; DB 105; Length 851;
Best Local Similarity	99.7%; Pred. No. 1.6e-301;
Matches	685; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1720	taatcagtgatgcaggagtgctgctactgaagcaatcacgaaaggatttcagtaagca 1779
Db 1	TAATCAGTGATGCGAGGAGTGTCTACTGGAAGCAATACGAAAGAGTATTTCAGCTACGCA 60
QY 1780	aagttagaagcagcgcgtgaacagagaagcaatgaacattgaaagcattgttccca 1839
Db 61	AAGTAGAAGAGCAGCGTGAACAGGAGCAATAGCATTGAACGATTTGAACAGATTTGCCA 120
QY 1840	ccatcctgtctgcgcgtattgctgttgaataatagctgattcggaagatgattcagaattg 1899
Db 121	CCATCCGTGTCTGCGCGTATTGCTGTCAATATAGTATTGCGAAGATGATTTCAGAAATTTG 180
QY 1900	atgaagtagattggttgagtagaagaaatgcatgataataattacaaactgaatgca 1959
Db 181	ATGAAGTAGATTGTTGGAGTAGAAGAAATGCATTGATAATATTACAAACTGAATGCA 240
QY 1960	aatgtcctttgtgctgttcttccctgaagaaattgttgctcattctagtgtttcttct 2019
Db 241	AATGTCTTTTGTGGTGTCTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTCTTCT 300
QY 2020	ttctctataataaatgaccttttctctccataaacttttcttcttaaggaataatag 2079
Db 301	TTTCTTTATAAATTAATGACCTTTTCTCTCATAACTTTTGATTTCTAAGGAAATATTAG 360
QY 2080	catactttcaactaaatgttttacagtggtctatcttttttttccctcctgaaagact 2139
Db 361	CATACATTTCAAACTAAATGTTTACAGTGGCTTATCTTTTTCCTCCCTGAAAGAACT 420
QY 2140	aatttgtcaataaac 2199
Db 421	AAATTGGTCAATAAATTAACCACTAAGTATTAGCATGGACGCTGTGTGTAGTAGTACGAT 480
QY 2200	tcagttttttgatatacttcttaattgtgtactttgtgtaatttttaatttaagaagcaact 2259
Db 481	TCAGTTTTTTTGATATATCTVAATTGTGTCTACTTTTGTGAATTTTAATTTAAAGAAAGCAACT 540
QY 2260	gaaattgaaatcttgaggcagctgttctactactaataatgagccttattccattctcgtagt 2319
Db 541	GAATTTGAAATCTTTGAGGCGAGCTGTACTACTAATGAGCCTTATTCCATTTCCTGATGT 600
QY 2320	tttaaaagaagaacacactgctgtgattatcgaataacacacacacacacacacacacacac 2379
Db 601	TTTAAAAAGAAAGAAACACTGCCTTGATTATACGAATACACTACAGAAAGTACATTAGCTTG 660

QY	904	tacataagcatattgaattgctaattggccaggcctctctaatgtttaaacaagaacctcaga	963
Df	290	TACATAAGCATATTGAATGCTTAATGGCCAGCCTCTCATTTTGAACAAGACCTTCAGA	349
QY	964	catacgtagatcatgatgatgacttactcaactttctgccttgccatttagtcagtga	1023
Df	350	CATACGTGGATCATATGATGATGATCTTACTCACTTTCGCCCTTGCCATTTAGTCAGATGA	409
QY	1024	gtgagcttctgactagagctgaggaaagggtattagtcagagaccacatgaaccacccacc	1083
Df	410	GTGAGCTTCTGACTAGAGCTGAGGAAGGGTATTAGTCAGACCACATGAACCACTCCAC	469
QY	1084	ctcaccaaatgcatgagcagagagatgaaaaaccatatccccacctgctatcagttctgta	1143
Df	470	CTCCACCAATGCATGAGCAGGAGATGCAAAACCGATACCCACCTGTATCACCTTCTGCTA	529
QY	1144	cagggttgatagaaaatcgccctcagtcaccagctcacaggcagacacacctgtgttgtga	1203
Df	530	CAGGTTTTGATAGAAAATGCCCTCAGTCACACAGCTACAGGCAGACACCTGTGTTTGTA	589
QY	1204	gcccaactccccacactctcccacacctcttccatctgccttgta	1250
Df	590	GCCCCACTCCCCCACCTCCTCCACCACTTTCCATCTGCCCTGTCA	636

RESULT	8
BG492540	
LOCUS	BG492540 821 bp mRNA EST 27-MAR-2001
DEFINITION	602536493Fl NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4655351 5', mRNA sequence.
ACCESSION	BG492540
VERSION	BG492540.1 GI:13454052
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 821)
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: CLONETECH Laboratories, Inc. DNA Sequencing by: Inyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM1444 row: j column: 24 High quality sequence stop: 726. Location/Qualifiers 1..821 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4655351" /clone_lib="NIH_MGC_59" /tissue_type="mucoepidermoid carcinoma" /lab_host="DH10B (TI phage-resistant)" /note="Organ; lung; Vector: pDNR-LIB (Clontech); Site_1: sfll (ggccgctcgcc); Site_2: sfll (ggccattatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCATTTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCAGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

FEATURES	source
----------	--------

BASE COUNT	247 a	149 c	153 g	272 t	
ORIGIN					
Query Match	17.6%	Score 461	DB 154	Length 821	
Best Local Similarity	100.0%	Pred. No. 4.1e-235			
Matches 461	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1690	agccaagcgccatccatcaacctacctgtgtaacagtgatgccaggagtggtgactggg	1749		
DB	1	AGCCAAAGGCCATCCATCAACCTACCTGTAATCAGTGTATCCAGGAGTGCTGCTACTGG	60		
QY	1750	aagcaatagaaaaaggtattcagctacgcacaaagtagaagacagcgtgaacaggaagcta	1809		
DB	61	AGCAATACGAAAGGTATTTCAGCTACGCAAAAGTAGAAGACGCGTGCAACAGGAAGCTA	120		
QY	1810	agcatgaagcattgaaacagatgttgccaccatcctgtctgcgcgtattgtctgtgaat	1869		
DB	121	AGCATGAACGCATTGAAACAGATGTTCACCATCTGTCTCGCGGTATTGCTGTTGAAT	180		
QY	1870	atagtattggaagatgattcagaatttgatgaagtagattggttgagtaagaaaaat	1929		
DB	181	ATAGTATGATTCGGAAGATGATTTCAGAAATTTGATGCAAGTAGATTGTTGGAGTAAGAAAAAT	240		
QY	1930	gcattgataaattacaaaactgaatgcaaatgctccttgggtgctgtgtctctgaaa	1989		
DB	241	GCATTGATAAATATTACAAACTGAATGCAATGTCCCTTGTGGTGCTGTGCTCTTGAAA	300		
QY	1990	atgttgggtcattctagtgtttgtcttttcttcttcttataataaatgaccttttctcc	2049		
DB	301	ATGTTTGGTCATTCTAGTCTTTTGGCTTTCTTTCTCTTATAATAAATGACCCCTTTTCCTCC	360		
QY	2050	ataactttgattctaaagaaaattagacatacttcaaacattcaaatgaatgtttcacagt	2109		
DB	361	ATACCTTTTGATTTCTAGAGAAATATTAGCATACTTTCAAACTAAATGTTTACAGTG	420		
QY	2110	gcttatctttttttccctcctgaaagacaaatttggtcaca	2150		
DB	421	GCTTATCTTTTTCCTCCCTGAAAGACTAATTTGGTCAA	461		
RESULT	9				
AI650710/c					
LOCUS	AI650710	462 bp	mRNA	EST	17-DEC-1999
DEFINITION	wb25g06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2306746	3'			
ACCESSION	AI650710				
VERSION	AI650710.1	GI:4734689			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 462)				
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.				
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.				
	cDNA Library Arrayed by: Greg Lennon, Ph.D.				
	DNA Sequencing by: Washington University Genome Sequencing Center				
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	www-bio.llnl.gov/dbirp/image/image.html				
	Insert Length: 971 Std Error: 0.00				
	Seq primer: -400P from Gibco				
	High quality sequence status: 438.				
	Location/Qualifiers				
FEATURES					
source	1..462				
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="IMAGE:2306746"					
/clone_lib="NCI_CGAP_GC6"					
/tissue_type="pooled germ cell tumors"					
/lab_host="DH10B"					
/note="vector: p773D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."					
BASE COUNT	182 a	73 c	64 g	143 t	
ORIGIN					
Query Match	17.5%	Score 460	DB 23	Length 462	
Best Local Similarity	100.0%	Pred. No. 1.4e-234			
Matches 460	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	2166	tlaagcagtcagcagctgtgttagagtagcagatcagtttttggatatactcttaattgt	2225		
DB	462	TTAAGCATGCGACAGCTGTTGTAGAGTAGCAGATTCAGTTTTTTTGATATATCTTAATTGT	403		
QY	2226	gtactttgtgaatttaatttaaagaaagcaactgaaatcctctgagggcagctgt	2285		
DB	402	GTACTTTGTGAATTTTAATTTAAAGAAAGCAACTGAAATTCGAAATCTTGAGGGCAGCTGT	343		
QY	2286	atctactaagccttattcattctcgtgatttttaaagaagaacaactgccttgat	2345		
DB	342	ATCTACTAATGAGCCTTATTCCTTCTCTGATGTTTAAAGAGAAACAACCTGCCTTGAT	283		
QY	2346	tatacgaatacacactcagaaagtagatttagctgttagttggaattcctctaaaggagt	2405		
DB	282	TATACGAATACACTCAGAAAGTACATTGTAGTGTGTAATTCCTCTAAAGGAATG	223		
QY	2406	cittgaaatttttcattattgttttatgtttttatactgctcatttggaattgttag	2465		
DB	222	CTTGAATTTTTCATATTGTTTATGTTTATATACCTGCTTATTTGAATGTTTAG	163		
QY					

M.J., Soares.F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-HT0230-040
500-110-f02&t3=2000-05-04&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 517.

FEATURES

source
1..555
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0230"
/dev_stage="Adult"
/note="organ: head_neck; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 180 a 111 c 110 g 154 t

Query Match 16.5%; Score 432; DB 164; Length 555;
Best Local Similarity 99.8%; Pred. No. 1.4e-219;
Matches 482; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 245 gccgtagtgaagaacacatcctaggaactgtgcccacagcactgctgaggg 304
DB 36 GCCGCTAGTGAAGAAGAACATCGATCCTAGGCACCTGTGCCACACAGCCTGCTAGG 95
QY 305 cattaagaatgaactgaatgtgaacaaatttctctggcaaatataattagacaact 364
DB 96 CATTAGAATGAATGGAATGTGTAACCAATATTTCTTGGCAATATATATTAGACAAC 155
QY 365 aagtagcctaagttaatatgctgaagatatatttggagaattattcaatgaagcacatag 424
DB 156 AAGTAGCCTAAGTAATATGCTGAAGATATATTTGGAGAAATTTATCAATGAAGCACATAG 215
QY 425 ttittcttcagatgaactcaatgaagaacgtgtgacgctttatctgttaggttac 484
DB 216 TTTTCCCTTCAGAGTCAACTCATTTGCAAGAACGCTGTGGACCGTTTATCTGTAGTGTAC 275
QY 485 acagcttgatccaaagaagaagaattgtcttggcaagatatcaaatgaggaagcttt 544
DB 276 ACAGCTTGATCCAAAGGAAGCAAGATTTGCTTTGCAAGATATACAAATGAGGAAGCTTT 335
QY 545 ccgaagtctacattcaagaccagcagcttttcgatcgcaagaactttgctctattccatt 604
DB 336 CCGAAGTTCTACAAATCAAGCAGCAGCTTTTCGATCGCAAGACTTTTGCTTATTCATT 395
QY 605 acaggagacatagatgttggtaacagcctccacctctcaatatactactcactcttag 664
DB 396 ACAGGAGACATAGATGTTTGTGAACAGCCTCCACCTCTCAATATACTCCTCTTAG 455
QY 665 agatgatggtaagaagctggaagttttataccaactcctctgattcttcttgaatctatg 724
DB 456 AGATGATGGTAAGAAGGCTGGAAGTTTTATACCAATCCTTCGTATTTCTTTGATCATG 515

QY 725 gaa 727
DB 516 GAA 518

RESULT 11

AW664122/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW664122 430 bp mRNA EST 06-APR-2000
h104e06.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2971330 3',
mRNA sequence.
AW664122
AW664122.1 GI:7456662
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 430)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: polyt not found
Seq primer: -40UP from Gibco
High quality sequence stop: 394.

FEATURES

source
1..430
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2971330"
/clone_lib="NCI_CGAP_GU1"
/tissue_type="2 pooled high-grade transitional cell
tumors"
/lab_host="DH10B"
/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
Site_1: Sall; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Library constructed by Life
Technologies."
BASE COUNT 173 a 64 c 59 g 134 t

Query Match 16.4%; Score 430; DB 119; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.6e-218;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2196 agattcagtttttgatataatcttaattgtgactttgtgaattttaataagaagc 2255
DB 430 AGATTCACTTTTGTATATATCTTAATTGTGTACTTTGTGAATTTTAATTAAGAAGC 371
QY 2256 aactgaaattgaaattcttgaggcagctgtatctactaataatgagccttattccattcctg 2315
DB 370 AACTGAAATTTGAAATCTTGAGGCAGCTGTATCTACTAATGAGCTTATTCCATTTCTCTG 311
QY 2316 atgttttaagaagaacacactgccttgattatatacgaatacactcagaagaagacattag 2375
DB 310 ATGTTTTAAAAGAAGAACAACACTGCCCTTGTATATACGAATACACTCAGAAAAGTACATTAG 251
QY 2376 ctgtgagttgtaattctcttaagaaggaatgcttgaatttttttatttattgttttttattgtt 2435
DB 250 CTGTGATGTTGAATTTCTCTTAAGGAATGCTTGAATTTTTCATTTATTTGTTTATTGTT 191

```

Qy 2436 ttatatacttgcccttatttgaaagttagcagtagtccccctccacttatatatattgtg 2495
Db 190 TTTATATACTTGCCCTATTGTAATGTTAGCAGTATCCCCCTCCACATTATATATTGTG 131
Qy 2496 gatagtatttgctgctataggaagttaaaaaactttccatggaatactctgactta 2555
Db 130 GATATGATTTTCTTGCCCTATAGGAGTTAAAAAAGCTTTTCCCATGTGAAATACTCTGACTTA 71
Qy 2556 aacatcatgtaactatcatcaactgtttaagaataaacagctgatttaataaagtgtcat 2615
Db 70 AACATACATGTAACATTCATCACTGTTAAGAATAACAGCTGATTTAAATAAGTGTTCAT 11
Qy 2616 tttaaaagtt 2625
Db 10 TTTAAAGTT 1

RESULT 12
AUI59926/c
LOCUS AUI59926 565 bp mRNA EST 25-OCT-2000
DEFINITION AUI59926 Y79AA1 Homo sapiens cDNA clone Y79AA1000685 3', mRNA
sequence.
ACCESSION AUI59926
VERSION AUI59926.1 GI:11021447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 565)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
,S. and Isogai,T.).
TITLE HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
,Y., Nagai,T., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
location/Qualifiers
1..565
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Y79AA1000685"
/clone_lib="Y79AA1"
/cell_type="retinoblastoma"
/cell_line="Y79"
/note="vector: pME18SFL3"
BASE COUNT 220 a 87 c 87 g 165 t 6 others
ORIGIN
Query Match 15.2%; Score 398; DB 108; Length 565;
Best Local Similarity 100.0%; Pred. No. 2.2e-201;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2228 acttgtgaatttttaataagaagaactgaactgaattgaatttgaggcagctgtat 2287
Db 398 ACTTTGTGAATTTTAAATTTAAAGAAAGCACTGAATTTGAATCTTGAGGCGAGCTGTAT 339
Qy 2288 ctactaatgacctattccattccctgattgttttaaaagaagaacacacgccttgatta 2347
Db 338 CTACTAATGAGCCCTATTCCCATTTCCGTATGTTTAAAGAAAGAAACACTGCGCTTGATTA 279

```

```

Qy 2348 tacgaatacactcagaagtagcatttagcttgtagttgtaattctctctaaagggaatgct 2407
Db 278 TAGCAATACACTCAGAAAGTACATTTAGCTTGTAGTGTGAATCTCTCTAAAGGAATGCT 219
Qy 2408 tgaatttttctcattatgttttattgtttttatatatactgctcatttgaatgtttaagca 2467
Db 218 TGAATTTTTTTCATTATTGTTTTTATTGTTTTTATATACATTCGCTTATTATTGAATGTTAGCA 159
Qy 2468 gtatcccttcccacttatattgtgtgatgatgattttgttcctatagagtaaaaa 2527
Db 158 GTATCCCTTCCCCTCTATATATTGTTGTGATGATGATTTCCTTCCTATAGAGTTAAAA 99
Qy 2528 actttccatgtgaataactctgactctgacttaaacatcatgtaacttaacataactgttaagaa 2587
Db 98 ACTTTTCCATGTGAATACTCTGACTTAAACATACATGTAACATTAACATGTTTAAGAA 39
Qy 2588 taacagctctgatttaataataatggttcatttttaaaagtt 2625
Db 38 TAACAGCTCTGATTTAATAAATGGTTTCATTTTAAAGTT 1

RESULT 13
BE731111
LOCUS BE731111 727 bp mRNA EST 15-SEP-2000
DEFINITION BE731111 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3841135 5',
mRNA sequence.
ACCESSION BE731111
VERSION BE731111.1 GI:10145091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 727)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLML at: image.llnl.gov
plate: LLCM532 row: e column: 08
High quality sequence stop: 669.
FEATURES
Location/Qualifiers
1..727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3841135"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 210 a 170 c 177 g 170 t
ORIGIN
Query Match 15.2%; Score 398; DB 139; Length 727;
Best Local Similarity 99.8%; Pred. No. 2.3e-201;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 214 ggtgaactggcacaagaagttaactcaagatgccctagtaggaagaacacatcgatacta 273

```

```
Db 157 GGTGAAGTGGCACAAGGTTAATCTCAAGATGCCGCTAGTAGGAAAGAAACATCGATCCTA 216
QY 274 ggcaattgtgcccacacagcactgcctagaggcaattagaatgaactgggaattgtgtaacca 333
Db 217 GGCACCTGTGTGCCACACAGCAGCTGCCTAGAGGCATTAAAGAAATGAACCTGGAATGTGTAACCA 276
QY 334 atatttccttgccaataataattagacaactaaagtagcctaaagtaataatgctgaagata 393
Db 277 ATATTTCCTTGGCAATATATAATTAGACAACCTAAAGTAGCCCTAAGTAATAATATGCTGAAGATA 336
QY 394 tattggagaattattcaatgaagcacatagtttttccctcagagtcacactcaattgcaag 453
Db 337 TATTTGGAGATTAATCAATGAAGCACATAGTTTTTTCCTTCAAGATCAACTCAATTCGAAG 396
QY 454 aacgtgtgaccgtttatctgttagttgttacacagctgtgacaaagggaagaattgt 513
Db 397 AACGTGTGGACCGTTTATCTGTGTAGTTTACACAGCTTGATCCAAAGGAAGAAATGGT 456
QY 514 ctttgaagataatacaatgagaaagccttcggaagttctacaaatcgaagaccagcagc 573
Db 457 CTTTGAAGATATAACAATGAGGAAAGCTTCCGAAAGTTCTACAATTCAGACCCAGCAGC 516
QY 574 ttttcacgcgaagacttgcttccattccattacagagacgtacgatgtttgtgaacagc 633
Db 517 TTTTCGATCCCAAGACTTTGCCATATCCATACAGGAGCTACGATGTTTGTGAACAGC 576
QY 634 ctcacactcctcaatatactactcactccttat 662
Db 577 CTCACCTCTCAATATACTACACTCCTTAT 605

RESULT 14
N95702 443 bp mRNA EST 09-APR-1996
LOCUS yv60c07.r1 Soares_multiple.sclerosis_2NBHMP Homo sapiens CDNA
DEFINITION clone IMAGE:277932 5', mRNA sequence.
ACCESSION N95702
VERSION N95702.1 GI:1268001
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 443)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 368.
Location/Qualifiers
1. .443
/organism="Homo sapiens"
/db_xref="GDB:3896308"
/db_xref="taxon:9606"
/clone="IMAGE:277932"
/clone_lib="Soares_multiple.sclerosis_2NBHMP"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 45"
/lab_host="PH10B (ampicillin resistant)"
/note="Vector: pT7FD (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
```

```
; 1st strand cDNA was primed with a Not I - oligo(drf)
primer [5',
TCTTACCAATCTGAAGTGGAGCGGCCGCAATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7n3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).
BASE COUNT 132 a 63 c 82 g 165 t 1 others
ORIGIN
Query Match 14.9%; Score 391; DB 187; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.2e-197;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 tccgtctgcgcgtattgtctgtgaatatagtgatcgcggaagatgattcagaatttgatg 1902
Db 2 TCCTGTCTCGCCGTATTGCTGTGTAATATAGTATGATTCGGAAGATGATTCAGAAATTTGATG 61
QY 1903 aagtagattggttgaggagtaagaaaaatgattgataataattacaaaactgaatgaaaa 1962
Db 62 AAGTAGATTGGTTGGAGTAAGAAAAATGCATTGATAAATATTACAAACATGAATGCAAAAT 121
QY 1963 gtccctgtggtgctgttctccttgaaaaatgttggctcattctctagttgttgcctcttt 2022
Db 122 GUCCTTGTGGTGCTGTGTCCTTGAAAAAGTTTGGTCATCTCTAGTGTGTTTTCCTTTCTTTT 181
QY 2023 ccttaataaataagacccttttctcctcaataaacttttgattctaaaggaaaaattagcat 2082
Db 182 CCTATAATAAATAGACCTTTTCTCTCCATAAATTTGATTCTAAGGAAAAATATTAGCAT 241
QY 2083 acatttcaaaactaaatgttttacagtggttattcttttttttttttttttttttttttttt 2142
Db 242 ACATTTCAAACTAAATGTTTACAGTGGCTTATCTTTTTCCTCCCTGAAAGACATAAT 301
QY 2143 ttggtcaataaaaccactaagattatgaacatgacagctgtgttagagtagcagattca 2202
Db 302 TTGCTCAATAAACCACTACTAGTATTAAAGCATGGACAGCTCTTGTAGAGTAGCAGATTCA 361
QY 2203 gtttttgatatatcttaattgtgtactttg 2233
Db 362 GTTTTGTGATATATCTTAATGTGTACTTTG 392

RESULT 15
AI223407/c 481 bp mRNA EST 29-NOV-1998
LOCUS qg48g10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1838466
DEFINITION 3', similar to SW:Y269_HUMAN Q92558 HYPOTHETICAL PROLINE-RICH
PROTEIN KIAA0269. ; mRNA sequence.
ACCESSION AI223407
VERSION AI223407.1 GI:3805610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 481)
REFERENCE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs@rmail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 719 Std Error: 0.00
Seq primer: 40UP from Gibco
High quality sequence stop: 326.
Location/Qualifiers
1. .481

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1838466"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	116 a	92 c	135 g	138 t
ORIGIN				
Query Match	14.8%	Score 389;	DB 17;	Length 481;
Best Local Similarity	99.8%	Pred. No. 1.4e-196;		
Matches 439;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 1516	cacccccaccacgcctctctgctccacctggcattcgaccatcatcacctgtcacag	1575		
Db				
QY 1576	ttacagctctgtctcatctcctctggtggtacatcccaactccatctactgccccaggtc	1635		
Db				
QY 1636	cccatgttccattaatgctccatctccatctccatcaagaattataacctgtcttgagccaa	1695		
Db				
QY 1696	agcgcctccatccaaacctaccctgtaacagtgatgccaggagtgctactggaagcaa	1755		
Db				
QY 1756	tacgaaaagggtattcagctacgcaaaagtagaagcagcgtgaacaggaagcgaagc	1815		
Db				
QY 1816	aacgcattgaaaacgattgtccaccatcctgtctcgcggtatgtctgtgaataatagtg	1875		
Db				
QY 1876	attcgggaagatgattcagaatttgatgaagtagattggtgagtaagaagaaatgcatg	1935		
Db				
QY 1936	ataaatattacaaaactgaa	1955		
Db				
QY 1955	ataaatattacaaaactgaa	1955		
Db				

Search completed: June 20, 2001, 22:22:15
Job time: 17966 sec

Result	Query No.	Score	Query		Length	DB	ID	Description
			Match					
C	1	352.8	13.4		359	14	Q61275	Human brain Express
C	2	245.2	9.3		359	21	A43416	Rat secreted expre
C	3	206	7.8		242	21	A42628	Human secreted exp
C	4	110.6	4.2		799	19	V55931	Nucleotide sequenc
C	5	110.6	4.2		1926	21	A50254	Epstein Barr virus
C	6	110.6	4.2		2580	21	A75454	Nucleotide sequenc
C	7	110.6	4.2		5452	20	X09023	Anti-sense strand
C	8	110.6	4.2		8705	20	Z23778	Vector pShuttle DN
C	9	110.6	4.2		9600	19	V21683	Vector plasmid pCM
C	10	110.6	4.2		10380	20	Z22248	Nucleotide sequenc
C	11	110.6	4.2		10596	14	O51731	Plasmid pcISBON f

Query Match	7.8%;	Score 206;	DB 21;	Length 242;
Best Local Similarity	98.6%;	Pred. No. 2.4e-39;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 3;	Gaps

Qy	739	tgaagatacacagaggataacagagaagaaaaaggaagcagacagacagaaaaatcttagatc	798
Db	21	tgaagatacacagaggataacagagaagaaaaaggaagcagacagacagaaaaatcttagatc	77
Qy	799	gtctcatgaaccagaaaaagtgcgaagcaccctcatgacagcgcgagaaatggcaga	858
Db	78	gtctcatgaaccagaaaaagtgcgaagcaccctcatgacagcgcgagaaatggcaga	137
Qy	859	agctgccccaaagtccagagctggctgaagatgatgctaattctctacataaagcatattg	918
Db	138	agctgccccaaagtccagagctggctgaagatgatgctaattctctacataaagcatattg	197
Qy	919	aagttgctaagtccccagcctctcattttgaaaacagacctc	960
Db	198	aagttgctaagtccccagcctctcattttgaaaacagacctc	239

RESULT	4
V55831/C	
ID	V55831 standard; DNA; 799 BP.
XX	AC
XX	AC
XX	XX
DT	18-NOV-1998 (first entry)
XX	
DE	Nucleotide sequence of the stabilising sequence-encoding insert.
XX	
KW	Fusion protein; stabilising polypeptide; proteolytic degradation;
KW	resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW	IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW	nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW	cancer; pathological condition; ss.
XX	
OS	Epstein-barr virus.
XX	
PN	WO9822577-A1.
XX	
PD	28-MAY-1998.
XX	
PF	17-NOV-1997; 97WO-IB01508.
XX	
XX	25-JUN-1997; 97US-0048945.
PR	15-NOV-1996; 96US-0030986.
XX	
PA	(MASU/) MASUCCI M G.
XX	
PI	Masucci MG;
XX	
DR	WPI; 1998-312463/27.
XX	
PT	New fusion proteins resistant to proteolytic degradation -
PT	comprising a core protein with a stabilising polypeptide comprising
PT	a peptide sequence containing glycine repeats
XX	
PS	Disclosure; Fig 4B; 120pp; English.
XX	
CC	This is a nucleotide sequence of the stabilising sequence-encoding
CC	insert. The invention provides a method for increasing the resistance
CC	of a core protein to proteolytic degradation that comprises linking
CC	inserting onto or into the core protein a stabilising polypeptide of
CC	formula [(Glya)x(Glyb)y(Glyc)z]n where Glya, Glyb, Glyc are 1-6
CC	sequential gly residues and x, y, z are Ala, Ser, Val, Ile, Leu, Met,
CC	Phe, pro or Thr and n can be anything between 1-66. x, y and z need n
CC	be identical from n repeat to n repeat. Alternatively a nucleic acid
CC	encoding the stabilising polypeptide can be linked onto or inserted
CC	into a nucleic acid encoding a core protein. The fusion proteins of the
CC	invention are more resistant to degradation by proteases and, thus, have
CC	a longer half-life than the unfused core protein. The products can be


```
XX 10-FEB-2000; 2000WO-US03819.
XX PF
XX 10-FEB-1999; 99US-0119571.
XX PR
XX 15-JUN-1999; 99US-0139172.
XX PA
XX (ELAN-) ELAN PHARM INC.
XX PI
XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;
XX Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
XX WPI; 2000-533011/48.
XX PT
XX Purified beta-secretase protein used in assays to discover inhibitors
XX PT which can be used for the treatment of amyloidogenic diseases e.g.
XX PT Alzheimer's disease
XX XX
XX Disclosure; Fig 13A-E; 121pp; English.
XX CC
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
XX beta-amyloid precursor protein to produce beta-amyloid peptide. This
XX enzyme is therefore implicated in the production of amyloid plaque
XX components which accumulate in the brains of individuals afflicted with
XX Alzheimer's disease. Inhibitors of beta-secretase are administered to
XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
XX disease-like pathology to test if they maintain or improve cognitive
XX ability or reduce the plaque burden. The compounds are used for the
XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
XX present sequence encodes a human beta-secretase enzyme.
XX SQ
XX Sequence 16080 BP; 3627 A; 4556 C; 3962 G; 3913 T; 22 other;

Query Match 4.2%; Score 110.6; DB 21; Length 16080;
Best Local Similarity 50.7%; Pred. No. 6.4e-16;
Matches 266; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

Qy 1163 ccctcagtcacacagctacagcagagacacctgtgttggagccacacccacacctcc 1222
Dy 10582 ccctcagtcacacagctacagcagagacacctgtgttggagccacacccacacctcc 10641
Qy 1223 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1282
Dy 10642 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 10701
Qy 1283 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1342
Dy 10702 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 10761
Qy 1343 agtaccacacacacacacacacacacacacacacacacacacacacacacacacacac 1402
Dy 10762 agtaccacacacacacacacacacacacacacacacacacacacacacacacacacac 10821
Qy 1403 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1462
Dy 10822 tccacacacacacacacacacacacacacacacacacacacacacacacacacacac 10881
Qy 1463 tgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 1522
Dy 10882 tgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 10941
Qy 1523 accacacacacacacacacacacacacacacacacacacacacacacacacacacac 1582
Dy 10942 accacacacacacacacacacacacacacacacacacacacacacacacacacacac 11001
Qy 1583 tctgcacacacacacacacacacacacacacacacacacacacacacacacacacac 1642
Dy 11002 tctgcacacacacacacacacacacacacacacacacacacacacacacacacacac 11061
Qy 1643 tccattaatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1687
Dy 11062 tccattaatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 11106
```

RESULT 15

```
C17657
ID C17657 standard; cDNA; 165 BP.
XX AC
XX C17657;
XX DT
XX 06-OCT-2000 (first entry)
XX DE
XX Human secreted protein 5' EST, SEQ ID NO: 21732.
XX KW
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX OS
XX Homo sapiens.
XX PN
XX EP1033401-A2.
XX PD
XX 06-SEP-2000.
XX PF
XX 21-FEB-2000; 2000EP-0200610.
XX PR
XX 26-FEB-1999; 99US-0122487.
XX (GEST ) GENSET.
XX PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX PT
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures
XX Claim 1; SEQ ID 21732; 71pp + CD-ROM; English.
XX CC
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX SQ
XX Sequence 165 BP; 43 A; 48 C; 25 G; 49 T; 0 other;
```

```
Query Match 4.2%; Score 110.4; DB 21; Length 165;
Best Local Similarity 99.1%; Pred. No. 7.3e-17;
Matches 111; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 cttcttgcacttgcgagtgatgaactgaactgaactgaactgaactgaactgaactgaact 60
Dy 54 cttcttgcacttgcgagtgatgaactgaactgaactgaactgaactgaactgaactgaact 113
Qy 61 caacattcacgtcctgcctataaccgattatgaattgattgattgattgattgattgattg 112
Dy 114 caacattcacgtcctgcctataaccgattatgaattgattgattgattgattgattgattg 165
```

Search completed: June 20, 2001, 17:22:48
Job time: 8300 sec

[illegible]

```

RESULT 12
US-08-757-669A-20
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:

```

[illegible][illegible]

RESULT 13
 US-09-010-928B-1/c
 ; Sequence 1, Application US/09010928B
 ; Patent No. 5994099
 ; GENERAL INFORMATION:
 ; APPLICANT: Lewis, Randolph V
 ; APPLICANT: Hayashi, Cheryl Y
 ; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
 ; TITLE OF INVENTION: CODING THEREFOR
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
 ; STREET: 8110 GATEHOUSE RD. SUITE 500E
 ; CITY: FALLS CHURCH
 ; STATE: VIRGINIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 22042
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/010,928B
 ; FILING DATE: 22-JAN-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Murphy Jr., Gerald M
 ; REGISTRATION NUMBER: 28977
 ; REFERENCE/DOCKET NUMBER: 1447-109P
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2830 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; FEATURE:
 ; NAME/KEY: -
 ; LOCATION: 1..2830
 ; OTHER INFORMATION: /note= "Flagelliform DNA sequence
 ; OTHER INFORMATION: taken from the 5' region. The putative star
 ; OTHER INFORMATION: position 219"
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 219..2830
 ; US-09-010-928B-1

Query Match
3.5% Score 93 DB 2: Length 2830;

1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 15:04:28 ; Search time 1943.75 Seconds
(without alignments)
12765.911 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1-cttctctgcacttgcggat.....aatgggtcattttaaaagtt 2625

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_estl1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estini:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estomi:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estrol1:*
83: em_estrol2:*
84: em_estrol3:*
85: em_estrol4:*
86: em_estrol5:*
87: em_estrol6:*
88: em_estrol7:*
89: em_estrol8:*
90: em_estrol9:*
91: em_estrol10:*
92: em_estrol11:*
93: em_estrol12:*
94: em_estrol13:*
95: em_estrol14:*
96: em_estrol15:*
97: em_estrol16:*
98: em_estrol17:*
99: em_estrol18:*
100: em_estrol19:*
101: em_estrol20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	844.6	32.2	851	105	AL526626	AL526626 AL526626
2	705	26.9	898	105	AL526602	AL526602 AL526602
3	664.2	25.3	700	108	AL142666	AU142666 AU142666
4	640	24.4	718	115	AW379048	AW379048 RC3-HT023
5	618.4	23.6	821	154	BG492540	BG492540 602536493
6	612.8	23.3	636	122	AW633370	AW633370 EST375443
7	608.6	23.2	988	173	BG111973	BG111973 602281809
8	605.2	23.1	650	175	BG288785	BG288785 602388115
9	565.8	21.6	652	122	AW956469	AW956469 EST368539
10	528.4	20.1	540	103	AI928895	AI928895 au65f06.x
11	511.2	19.5	727	139	BE731111	BE731111 601566327
12	510.6	19.5	835	174	BG176906	BG176906 602313474
13	496	18.9	497	111	AW071606	AW071606 wt94a11.x
14	495.4	18.9	555	164	BE148072	BE148072 RC3-HT023
15	493.6	18.8	547	32	AV664100	AV664100 AV664100
16	488.4	18.6	565	108	AU159926	AU159926 AU159926
17	478.8	18.2	1142	174	BG168604	BG168604 602344810
18	478	18.2	492	136	BE463994	BE463994 hy19d07.x
19	463.6	17.7	481	17	AI223407	AI223407 qg48g10.x
20	460	17.5	462	23	AI650710	AI650710 wd25906.x
21	455.8	17.4	878	153	BG426480	BG426480 602492882
22	449.8	17.1	505	151	BF651492	BF651492 274285 MA
23	435.8	16.6	451	13	AA885055	AA885055 am11c01.s
24	433.2	16.5	778	140	BE787088	BE787088 601476633
25	430	16.4	430	119	AW664122	AW664122 h104e06.x
26	426.4	16.2	894	168	BF700682	BF700682 602128670
27	423.2	16.1	512	170	BF830660	BF830660 RC3-HT023
28	421.8	16.1	443	187	N95702	N95702 YV60c07.r1
29	418.2	15.9	888	169	BF796356	BF796356 602258682
30	405.4	15.4	442	159	N59851	N59851 yz32a12.s1
31	405.4	15.4	600	103	AI929316	AI929316 au65f06.y
32	402.8	15.3	406	12	AA846712	AA846712 a141c05.s
33	399.6	15.2	764	169	BF785010	BF785010 602110960
34	395.8	15.1	415	118	AW572884	AW572884 hf17a04.x
35	391.4	14.9	692	105	AW079676	AW079676 DKF2p434B
36	389.8	14.8	863	172	BG028914	BG028914 602293024
37	387	14.7	394	24	AI750422	AI750422 ch02g10.y
38	384.4	14.6	395	15	AI061122	AI061122 an35e08.x
39	383	14.6	388	119	AW673234	AW673234 ba64c12.x
40	377.4	14.4	476	151	BF651493	BF651493 274286 MA
41	374.8	14.3	394	159	N63410	N63410 YV60c07.s1
42	362	13.8	390	8	AA515859	AA515859 n130303.s
43	361	13.8	640	145	BF143773	BF143773 601789661
44	356.8	13.6	427	16	AI154223	AI154223 ud30d06.r
45	355.6	13.5	441	112	AW140599	AW140599 EST290590

ALIGNMENTS

RESULT	1
AL526626	
LOCUS	AL526626 851 bp mRNA EST 13-FEB-2001
DEFINITION	AL526626 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC020Y018 5 prime, mRNA sequence.
ACCESSION	AL526626
VERSION	AL526626.1 GI:12790119
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers	Source
1..851	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="CS0DC020Y018"	
/clone_lib="LTI_NFL003_NBC3"	
/sex="male"	
/tissue_type="neuroblastoma cells"	
/lab_host="DH10B"	
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	
BASE COUNT 263 a 132 c 153 g 299 t	4 others
ORIGIN	
Query Match 32.2%; Score 844.6; DB 105; Length 851;	
Best Local Similarity 99.2%; Pred. No. 9.1e-188;	
Matches 844; Conservative 4; Mismatches 3; Indels 0; Gaps 0;	

QY	1720	taatcagtgatgcagagtgctactggaagcaatacgaagaaggtattcagctagca	1779
Db	1	TAATAGTAGTATGTCAGAGGTGCTACTTGGAGCAATACGAAAGGTATTACGAC	60
QY	1780	aatgaagagcagcgtgaacagaaagcaagcattgaaacagatgttgcca	1839
Db	61	AAGTAGAGAGCAGCGTGAACAGCAAGCAATAGCATGCAAGCAATGGAAGTTGCCA	120
QY	1840	ccatcctgtctccgctattgtgtggaatagtagtattcggaagatgattcagaattg	1899
Db	121	CCATCCTGTCTCGCGTATTGCTGTTCAATATAGTATCGGAAGATGATTACGAATTG	180
QY	1900	atgaagtagattgggtggaagaaataatgcatgataataattacaaactgaatgca	1959
Db	181	ATGAAGTAGATTGGTGGAGTAGAAGAAATGCATTGATAAATATTACAAACTGAATGCA	240
QY	1960	aatgtcctttgtggtcttcttctgaagaattgttgctattcattctagtttcttct	2019
Db	241	AATGTCTCTTTGTGGTGTCTTGTCCCTTGAAATGTTTGTGCTATTCTAGTGTTCCTTCT	300
QY	2020	tttcctataataaataaaccttttccataacttttgatttcttaaggaaataatag	2079
Db	301	TTTCCCTTAAATTAATGACCCCTTTTCCTCCATAACTTTTGTATTTAAGGAAATATAG	360
QY	2080	catacatctcaactaaatgttttacagtggtcttatcttttttccccctgaaagact	2139
Db	361	CATACATTTCAAACTAAATGTTTACAGTGCTTATCTTTTTTTTCCCTGAAAGACT	420
QY	2140	aattggctcaataaacactaagatttaagcatggacagctgtgttagatgagat	2199
Db	421	AATTTGGTCAAAATAAACCACTAAGTATTAAAGCATGGACAGCTGTTGTAGAGTAGCAGAT	480
QY	2200	tcagtttttgatatacttaattgttactttgtgaatttttaatttaagaaagcaact	2259
Db	481	TCAGTTTTTTTGATATATCTTAATTGTGTACTTTTGTGAATTTTAAATTTAAGAAAGCAACT	540
QY	2260	gaattgaattcttgaggcagctgttactactaatgagccttattccatttctctgattg	2319
Db	541	GAAATTGAATCTTGAGGCGAGCTGTATCTACTTAATGAGCCTTATTCCTTCCTGATGT	600
QY	2320	tttaagaagaacactcgtctgattatcagatacactcaagaagaattacatttagcttg	2379
Db	601	TTTAAAAAGAAACACTGCCTTGTATTATACGATACTACAGAAAGTACATTTAGCTTG	660

Db 421 GATCTATGGAAGAAAAATGTTGCAAGATACAGAGGATAAGAGGAGGAGGAG 362
Oy 777 cagaagcagaaaaatctagatcgctctcatcaaacagaaaaagtgccaagcacctcat 836
Db 361 CAGAAAGCAGAAAAATCTAGATCGTCTCATGAACACAGAAAAAGTGCCAGAGCACCTCAT 302
Oy 837 gacagggcgagaaatggcagaagctggcccaagctcagagctggctgaagatgatgct 896
Db 301 GACAGCGCGGAGATGCGAGAGCTGCGCCCAAGTCCAGAGCTGCGCTGAAGATGATGCT 242
Oy 897 aatctcttaccataagcatatgaagtgtgtaagtggccagcctctcattttgaaacaaga 956
Db 241 AATCTCTTACATAAGCATATGAAGTTGCTAATGCGCCAGCGCTCTCATTTTGAACAAGA 182
Oy 957 cctcagacatacgtgatcatatgatgatgattctactcaatttctgccttgccatttga 1016
Db 181 CCTCAGACATAGTGGATCATATGATGGATGGATTTACTCACCTTTCTGCGCTTGCATTTAGT 122
Oy 1017 cagatgagtgagctctgactagagctgaggaagggttatagtcagaccacatgaacca 1076
Db 121 CAGATGAGTGAGCTTCTGACTAGAGCTCAGGAAAGGGTATTAGTCAGACCACATGAACCA 62
Oy 1077 cctccacctccaccaatgatggagcaggagatgcaaacccgataccacacctgtatcagt 1136
Db 61 CCTCCACCTCCACCAATGCATG-----ACGGAGATCAAAACCGATCCGCGCTGTATCAGT 7
Oy 1137 tctgct 1142
Db 6 TCTGCT 1

RESULT 5
BG492540 BG492540 821 bp mRNA EST 27-MAR-2001
LOCUS 602536493P1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4655351 5',
DEFINITION mRNA sequence.
ACCESSION BG492540
VERSION BG492540.1 GI:13454052
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 821)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM1444 row: j column: 24
High quality sequence stop: 726.
Location/Qualifiers
1. 821
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:4655351"
/clone_lib="NIH_MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatgcc);
Double stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dT(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 247 a 149 c 153 g 272 t
ORIGIN

Query Match 23.6%; Score 618.4; DB 154; Length 821;
Best Local Similarity 93.1%; Pred. No. 1.3e-134;
Matches 747; Conservative 0; Mismatches 41; Indels 14; Gaps 9;

Oy 1690 agcaaaagcgcacccatccatcaacccctacctgtaactcagtgatgccagagtgctgactgg 1749
Db 1 AGCCAAAGCGCCATCCATCAACCCCTACCTGTATYACGTGATGCCAGAGTGTGCTACTGG 60
Oy 1750 aagcaatacgaagaaggtattcagctacgcgaagtagaagagcagcgtgaaacaggaagcta 1809
Db 61 AAGCAATACGAAAGGTATTTCAGCTACGCAAGTAGAAGAGCAGCGTCAACAGGAAGCTA 120
Oy 1810 agcatgaacgattgaaacgattgtgccacacatcctctctcgcgtatttgcgttgat 1869
Db 121 AGCATGAACGCAATGAAACGATGTGCCACCATCTCTCGCGGTATGCTGTGAAT 180
Oy 1870 atagtattcggaagatgattcagaatttgatgaagtagattgttgagtaagaagaaat 1929
Db 181 ATAGTGATTCGGAAGATGATTCAGAAATTTGATGAAGTAGATTGGTTGGAGTAAGAAAAAT 240
Oy 1930 gcattgataataattacaaaaactgaatgcaaatgctcctttgtggtgcttgccttgaaa 1989
Db 241 GCATTGATAAATATTACAAAACCTGAATGCAATGCTCTTGTGGTCTGTTCCTTGAAA 300
Oy 1990 atgttggctcatttagtattgtcttcttcttcttcttcttcttcttcttcttcttcttcc 2049
Db 301 ATGTTTGGTCAATCTAGTGTGTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 360
Oy 2050 ataacttttgattctcagaagaaatattagcatatcattcacaactaaactaaatgtttcacagt 2109
Db 361 ATACTTTTGAATTTCTAAGGAAATATTAGCATATCATTTCAAACTAAATGTTTACAGTG 420
Oy 2110 gcttatcttttttttccccctgaaaagactaatttgggtcaa-ataaacccactaagtatta 2168
Db 421 GCTTATCTTTTTTTTCCCTGCTGAAAAGACTAATTGGTCAACATAAACCACTAAGTATTA 480
Oy 2169 agcatggacgctgtttagtagtagcagattcag-ttttttgatatattcttaattgtgt 2227
Db 481 AGCATGGACAGCTGTTGTAGAGTAGCAGATTCAAGTTTTTTTGTATATATCTTAATTGTGT 540
Oy 2228 actttgtgaattttaattta--aagaaagcaactgaaattgaaactcttgagggcagctgt 2285
Db 541 ACTTTGTGAATTTTAAATTTACAGAAAGCAACCTGAAATTTGAAATCTTGAGGGCAGCTGT 600
Oy 2286 atctactaaatgagccttattcca-tttctctgattgtttt--aaaagaagaacactgcctt 2342
Db 601 ATCTACPAATGAGCCTTATTCCTCATTTCTCTGATGTTTTCACACAGAGAAGAACACTGCCT 660
Oy 2343 gattatacag-aatacactcagaaa---gtacatttagctgttagtgtaa-ttctctta 2397
Db 661 GATTATACGAAATACACTCAGAAACGTTACAGTTTAGCTTTTGTAGTGTGTAACCTTCTCTTA 720
Oy 2398 aaggaatgctgaatttttctatttattgttttattgttttttatttatttatttatttga 2457
Db 721 CAGGAATGCTGGAATTTCTTCTCATGG--TCCATTGTACAAATATACCTTCTTATTTCGA 778
Oy 2458 atgttagcagtatcccccttcc 2479
Db 779 CTGCTAGCAAGTATCCCTTCC 800

RESULT 6
AW963370

LOCUS AW963370 636 bp mRNA EST 01-JUN-2000
DEFINITION EST375443 MAGE resequences, MAGH Homo sapiens cDNA, mRNA sequence.
ACCESSION AW963370
VERSION AW963370.1 GI:8153206
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 636)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 187
Seq primer: Reverse.
FEATURES
source
1..636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGH"
/note="Vector: pBluescriptSKm"
BASE COUNT 207 a 125 c 140 g 164 t
ORIGIN
Query Match 23.3%; Score 612.8; DB 122; Length 636;
Best Local Similarity 98.7%; Pred. No. 2.5e-133;
Matches 628; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 451 agaacgtgtggaccgtttatctgttagttacacagcttgatccaaagggaagaat 510
DB 1 AAGACGCTGGACCGCTTTATCTGTAGTTACACAGCTTGATCCAAAGGAAGAAT 60
QY 511 tgccttgcagataatacaaatgagaaagctttccgaagttctacaaattcaagaccgc 570
DB 61 TGCTTTGCAAGATATACAAATGAGAAAGCTTTCGCAAGTTCTACAATTCAGACAGC 120
QY 571 agcttttcgatcgcagactttgcctattccattacagagacgtacgatttttgaaac 630
DB 121 AGCTTTTCGATCCCAAGACTTTGCCCTATTCCTATTCAGGAGACGTACGATTTTGGAAC 180
QY 631 agctccacctctcaataataactcactccttattagatgatggtaagaaggtctgaagt 690
DB 181 AGCTCCACCTCTCAATATACTACTCTCTATAGAGATGATGCTAAGAGGTCCTGAAGT 240
QY 691 ttataccatctcctcgattttttttgtatctatcgaaagaaaaattgtgcaagatacag 750
DB 241 TTATACCAATCTCGTATTTCTTGTATGATGATGAAAGAAAAAATGTTGCAAGATACAG 300
QY 751 agcataagaggaagaaagaggaagacagacagacagaaaaatctagatcgtcctcatgaac 810
DB 301 AGGATAGAGGAAGAAAGAGGAAGCAGAACGAGAAATATCTAGATCGTCCTCATGAAC 360
QY 811 cagaaaaagtgcgaagcacctcatgacagcgcgagaaatggcagaagctggcccaag 870
DB 361 CAGAAAAAGTGCACAGCACCTCATGACAGGCGGCGAGAAATGGCAGAGCTGGCCCAAG 420
QY 871 gtccagagctggctgaagatgtagtctaattctctacataagcatattgaagttgctaag 930
DB 421 GTCCAGAGCTGGCTGAAGATGATGCTAATCTCTACATAGCATATTTGAAGTTGCTAATG 480
QY 931 gccagacctctcattttgaaacaagacctcagacatacgtggatcatatgtgatgattctt 990
DB 481 GCCAGCCTCTCAATTTGAAACAAGACCTCAGACATACGTGGATCATATGATGGATCTT 540

QY 991 actcaattctgcttgccatttagtcagatgagctgagcttctgactagagctg-aggaag 1049
DB 541 ACTTACTTTTGGCTTCCATTAGTCAGATGAATGAGCTTTTGGACTAGAGCTGAAGAA 600
QY 1050 agggatttagtcagaccacatgaaccacacctcaccct 1085
DB 601 AGGGTATTAGTACAGACCACTTGAACCAACTTCACCT 636
RESULT 7
BG111973 988 bp mRNA EST 30-JAN-2001
LOCUS 602281809F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4369058
DEFINITION mRNA sequence.
ACCESSION BG111973
VERSION BG111973.1 GI:12605479
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 988)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10024 row: b column: 03
High quality sequence stop: 687.
FEATURES
source
1..988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4369058"
/clone_lib="NIH_MGC_86"
/tissue_type="osteosarcoma, cell line"
/lab_host="PH10B (phage-resistant)"
/note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NOTI; Site 2: Sali; Cloned unidirectionally; oligo-dT primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 313 a 224 c 203 g 248 t
ORIGIN

Query Match 23.2%; Score 608.6; DB 173; Length 988;
Best Local Similarity 94.1%; Pred. No. 2.8e-132;
Matches 699; Conservative 0; Mismatches 34; Indels 10; Gaps 6;

QY 414 gaagcacatgttttcccttcagatcactcattcaagaacgtgtgacccttatct 473
DB 1 GAAGACATAGTTTTTCTTTCAGAGTCACTCAATGCAAGACGTGGACCGTTATCT 60
QY 474 gttagtgttacacagcttgatccaaagggaagaatgtctttgcaagataaacaatg 533
DB 61 GTTAGTGTACACAGCTTGATCCAAAGGAAGAATTGCTTTGCAAGATATAACAATG 120
QY 534 agaaagctttccgaagttctacattcaagaccagcagctttcttcagctcgaagcttg 593
DB 121 AGAAAGCTTTCGAGAGTTCTACAATTCAAGACCAAGCAGCTTTTCGATCGCAAGACTTG 180
QY 594 cctattccattacagagacgtacgatgtttgtgaacagcctcaccctcctcaatactc 653
DB 181 CCTATTCCATTACAGGAGAGGTACGATGTTGTGGAACAGCTCCACCTCTCAATATACT 240

/note="organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 199 a 163 c 134 g 156 t

ORIGIN

Query Match 23.1% Score 605.2; DB 175; Length 652;

Best Local Similarity 97.3%; Pred. No. 1.5e-131;

Matches 638; Conservative 0; Mismatches 13; Indels 5; Gaps 2;

Qy 649 tactcactccttatagagatgatgttaagaaggtcttgaagtttttataccaatccttcgt 708

Db 1 TACTCCTCTTATAGAGATGATGGTAAGAAGGTCTGAAGTTTATACCAATCCTTCGT 60

Qy 709 attctttgatctatgaaagaaatgttgcaagatcacagagataagaggaagaaa 768

Db 61 ATTTCTTTGATCTATGGAAGAAAAATGTTGCAAGATACAGAGGATAAGAGGAAGAAA 120

Qy 769 agaggaagcagaagcagaaaaatctagatcctctcatgaaccagaaaaagtgcaag 828

Db 121 AGAGGAAGCAGAGCAAGAAAAATCTAGATCTGCTCATGAACCAAGTGGCAAGAG 180

Qy 829 cactcatgacagcgccgagaatggcagaagctggcccaaggtccagagctggctgaag 888

Db 181 CACTCATGACAGCGCGGAGAGTGGCAGAGCTGGCCCAAGTCCAGAGCTGGCTGAAG 240

Qy 889 atgatgtaattctttacataagcatattgaagttgttaattggtccagcctctcatttg 948

Db 241 ATGATGCTAATCTCTTACATAAGCATATTGAAGTTGCTTAATGGCCAGCCTCTCATTTTG 300

Qy 949 aacaaagacctcagacatacgtgagatcatatggatggatgttactcactttctgccttgc 1008

Db 301 AAACAAGACCTCAGACATACCTGGATGATATGATGGATGATCTTACTCATCTTCTGCTTGC 360

Qy 1009 catttagtcagatgagcttctgactagagctgaggaaggggtatttagtcagaccac 1068

Db 361 CATTTAGTCAGATGAGTGAGCTTCTGACTAGAGCTGAGGAAGGGGTATTAGTCAGACCAC 420

Qy 1069 atgaacacacctccacccacccaatgcatggagcagagatgcaaaacccgatacccaacct 1128

Db 421 ATGAACACACCTCCACCTCCACCAATGATGAGCAGGAGATGCAAAACCGATACCCACCT 480

Qy 1129 gtatcagttctgtacaggttttgatagaaaaatccctcagtcaccagctacagcagaa 1188

Db 481 GTATCAGTTCTGTACAGGTTTGATAGAAAAATCGCCCTCAGTCACCCAGCTACAGGCAGAA 540

Qy 1189 cactgtgtttgtgagcccccactccccacccctctccaccacctcttccatctgccttgc 1243

Db 541 CACTGTGTGTGTGAGCCCCCATCCCCACTCTCCACA----CTCTTCCATCTGCCTTGT 596

Qy 1249 caacttctcattaaagagcttcaatgacttcaactctccccctccctccagctaccctccc 1304

Db 597 CAACTTCTCTATTAAGAGCTTCAATGACTTTTCAATCCT-CCCTCCAGTAACTCCC 651

RESULT 9

AW956469

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 650)

Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt

,J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and

Quackenbush,J.

AW956469 650 bp mRNA EST 01-JUN-2000

EST368539 MAGC resequences, MAGD Homo sapiens cDNA, mRNA sequence.

AW956469

AW956469.1 GI:8146152

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 650)

Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt

,J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and

Quackenbush,J.

Qy 634 actccttagagatgagtgtaaaagagctgtaagttttataccaatccttcgtatttc 713

Db 241 ACTCTTTATAGAGATGATGTTAAAGAAGCTCTGAAGTTTATACCAATCCTTCGTATTTC 300

Qy 714 ttgatctatgaaagaaaaaatgttgaagatacagagggataagagaagaaagagg 773

Db 301 TTTGATCTATGGAAGAAAAATGTTGCAAGATACAGAGGATAAGAGGAAGGAAAGAGG 360

Qy 774 aagcagaagcagaaaaactta-gatcgctccctcatgaaccagaaaaagtccaagagacc 832

Db 361 AAGCAGAAGCAGAAAAATCTATGATCGTCTCATGAACCAAGAAAGTGCCAAGAGCACC 420

Qy 833 tcatgacagggcgagagaatggcagaagctgcccgaaggtccagagctggctgaagatga 892

Db 421 TCATGACAGCGCGGAGAGTGGCAGAACCTGG-CCAAGTCCAGAGCTGGCTGAAGATGA 479

Qy 893 tgcataatctttacataagcatattgaa---gttgctaatggccagcctctcatttga 949

Db 480 TGCTAATCTCTTACATAAGCATATTGAAGTTTGCTAAATTTGGCCAGCTCTCATTTTGA 539

Qy 950 aacagacccagacacatcgtgagatcatatggatggatcttactcactttctgc---tt 1006

Db 540 AACAAGACCTCAGACATACGTGGATCATATGGATGGATCTTACTCACCTTCTGCTTTTG 599

Qy 1007 gccatttagtcagatgagcttctgactagagctgaggaaggggtatttagtcagacc 1066

Db 600 CCAATTTAGTCAGATGAGTCTGACTAGAGCTGAGGAAGGGTATTAGTCAGACC 659

Qy 1067 acatgaacacccctccacccacccaatgcatggagcagagagatgcaaaacccgataccacc 1126

Db 660 CACATGAACA-CTCCAACCTCCACCAATCTTGGAGCAGGAGATGC-ATACGAATTCGCCAC 717

Qy 1127 ctgatacgttctgtacaggtt 1149

Db 718 CTGTATCAGTCTGCTACCGGTT 740

RESULT 8

BC288785

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BC288785 652 bp mRNA EST 21-FEB-2001

602388113F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4517163 5',

mRNA sequence.

BC288785

BC288785.1 GI:13043972

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 652)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM10409 row: m column: 04

High quality sequence stop: 638.

Location/Qualifiers

1. 652

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4517163"

/clone_lib="NIH_MGC_93"

/tissue_type="transitional cell papilloma, cell line"

/lab_host="DH10B (phage-resistant)"

FEATURES

source

EST 23-AUG-1999

Query Match
Best Local :
Matches 52%

ch
l Similarity
529; Conserv

QY 2276 ggcagcgtatctactaatgagccctattccatttcttgatgttttaaaagaagaaca 2335
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 GGGCAGCTGTGTC TACTAATGAGCCCTATTCCATTCTCATGTTTAAAGAAGAACCA 301


```
QY 2336 ctgccttgattacgaatacactcagaagtagcttgtagttagtgaattctct 2395
Db 300 CTGCTCTGATTATAGCAATACACTCAGAAAGTACATTAGCTGTGAGTTGAATCTCT 241
QY 2396 taaagaatgcttgaaatttttccattattgtttttatgtttttatatactgcttattt 2455
Db 240 TAAAGGAATGCTGTAATTTTTCATATTATTGTTTATGTTTATATACTGCTCTTATT 181
QY 2456 gaatttttagcagtagtccccctcccaacttatattgtgtgatagatttgcctgacct 2515
Db 180 GAATGTTTAGCAGTATCCCTTCCACACTATATATGTTGTGATATGATTTCCTTGCCTA 121
QY 2516 taggagttaaaaactttccatctgtaataactctgacttaaacatacatgtaacttacct 2575
Db 120 TAGGAGTTAAAACCTTTCCATGTAATACTCTGACTTAACATACATGTAACATTACAT 61
QY 2576 aactgttaagaataacagctgtgatttaataaagtgttcatattttaaagt 2625
Db 60 AACTGTTAAGATAACAGTCTGATTTAATAAATGTTTCATTTTAAAGTT 11

RESULT 11
BE731111 727 bp mRNA EST 15-SEP-2000
LOCUS 601566327F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3841135 5',
mrna sequence.
ACCESSION BE731111
VERSION BE731111.1 GI:10145091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 727)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Cloned through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM532 row: e column: 08
High quality sequence stop: 669.
Location/Qualifiers
1. 727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3841135"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Size-selected by
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
210 a 170 c 177 g 170 t

Query Match 19.5%; Score 511.2; DB 139; Length 727;
Best Local Similarity 94.7%; Pred. No. 2e-109;
Matches 551; Conservative 0; Mismatches 28; Indels 3; Gaps 2;
QY 204 aaattctcaaggtgaactgacacaaaggttaattctcaagtagtcgtagtgaagaagaac 263

```
Db 147 AAGATTTCACGCTGAACCTGGCACAAAGGTTAATCTCAAGATGCCGTAGTGAAGAAGAAC 206
QY 264 atcgatcttaggcacttggccacacagcactgctcctagagcattaaagaatgaactgaa 323
Db 207 ATCGATCTTAGGCACCTTGTGCCACACAGCAGCTGCTAGAGGCATTAAAGAATGAAGTGA 266
QY 324 tgtgaaccaataattctcttgccaaataataattagacaactaagtagcctgaagtaaatat 383
Db 267 TGTGTAAACCAATATTTCCTTGGCAAAATATATTAGACAACCTAAGTAGGCTTAAGTAATAT 326
QY 384 gctgaagatatatttgggagaattattcaatgaagcacatagtttttctcctcagagtcacac 443
Db 327 GCTGAAGATATATTGGAGAAATATTCAATGAAGCACATAGTTTTTCTTCCAGAGTCAAC 386
QY 444 tcattgcaagaacgtgtggaacgtttatctgttagtgttacacagcttgatcccaagaaga 503
Db 387 TCATTGCAAGAACGCTGTGACCGTTTATCTGTAGTTGTACACAGCTTCATCCCAAGAA 446
QY 504 gaagaattgctcttgcaagataatacaatgaggaagcgtttcccgaaagttctacaattcaa 563
Db 447 GAAGAACTGCTCTTTCGAAGATATAACAATGAGGAAGCTTTCCGGAAGTTCTACAATTCA 506
QY 564 gaccagcagcttttcgatcgcaagactttgctattcctcattccacagggagcgtacgatgtt 623
Db 507 GACCAGCAGCTTTTCGATCGCAAGACTTTTGCTATTTCATTACAGGAGACGTACGATGTT 566
QY 624 tgtgaacagcctccacctcctcaataatactactcctccttatagagatgagtggaagaagt 683
Db 567 TGTGAACAGCCTCCACCTCTCAATATATCTACTCTCTTATCGAGATGATGGTGAAGAAG 626
QY 684 ctgaagttttataccacatcctctgattcttcttctgctatctatgaaa--gaaaaaatgttgc 741
Db 627 CTGAAGTTTATACCAATCCCTCGTA-TTCITTGATCTATGGACACAGCAACAATGTTGC 685
QY 742 aagatacagagagataagaaggaaggaaggaaggaaggaaggaagc 783
Db 686 CCGATACAGAGGATCAACAGGCGCACGGCAACAGACAGCAGGAAGC 727

RESULT 12
BG176906 835 bp mRNA EST 06-FEB-2001
LOCUS 602313474F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419042 5',
mrna sequence.
ACCESSION BG176906
VERSION BG176906.1 GI:12683609
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 835)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Cloned through the I.M.A.G.E. Consortium/LLNL at:
found through the I.M.A.G.E. Consortium/LLNL at:
Plate: LLAM10154 row: d column: 19
High quality sequence stop: 695.
Location/Qualifiers
1. 835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4419042"
/clone_lib="NIH_MGC_85"
```

FEATURES
source

Db 257 ATTAGCTTGTAGTGTGAATTCCTTAAGGAATGCTTGAATTTTCATTAATTTGTTTT 198
 Qy 2430 attgtttttatattccttatttgaatttttagcagtagtacccttccaccattatata 2489
 Db 197 ATTGTTTTTATATACCTGCCCTTATTGAATGTTTAGCAGTAGTCCCTTCCCACTTATATA 138
 Qy 2490 ttgtgtgatgattttgtcttgcctataggaggtttaaacttttccatgtgaaatacactct 2549
 Db 137 TTGTGTGATAGTATTTGCTTGGCTATAGGAGTTAAACATTTTCCATGTGAATACTCT 78
 Qy 2550 gacttaacatacatatgttaactataactgtttaagaataacagctgtgatttaataatg 2609
 Db 77 GACTTAAACATACATGTAACCTTACATACTCTTAAGAATAACAGTCTGATTTAATAAATG 18
 Qy 2610 gttcatttttaaaagt 2625
 Db 17 GTTCAATTTTAAAGTT 2

RESULT 14
 BE148072 555 bp mRNA EST 21-JUN-2000
 LOCUS RC3-HR0230-040500-110-f02 HT0230 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE148072
 ACCESSION BE148072
 VERSION BE148072.1 GI:8610796
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 555)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC3-HT0230-040
 500-110-f02&t3-2000-05-04&t4-1)

Seq primer: puc 18 forward
 High quality sequence set: 517.
 Location/Qualifiers
 1. 555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR0230"
 /dev_stage="Adult"
 /note="Organ: head_neck; Vector: puc18; Site:1: Sma1;
 Site:2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

FEATURES
 source
 180 a 111 c 110 g 154 t

BASE COUNT
 ORIGIN
 Query Match 18.9%; Score 495.4; DB 164; Length 555;

Best Local Similarity 98.3%; Pred. No. 9.4e-106;
 Matches 532; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 Qy 212 aaggtgaactggcacaaggttaattctcaagatgccgctagtgaaagaacacatgcacc 271
 Db 5 ACGGTGTTCTGGCACAAAGGTT-ATCTCAAGA-GCCGCTAGTGAAGAAACATCGATCC 62
 Qy 272 tagcacttggccacacagcactgcctagagcattaaagtaagaaactgggaatgtgtaac 331
 Db 63 TAGCACATGTGGCCACACAGCAGCTCCCTAGAGGCAATTAAAGATGAACCTGGATGTGTAAAC 122
 Qy 332 caatatttcttggcacaataataattagacaactaagtagccttaagtaataatgtctgaaga 391
 Db 123 CAATATTTCCTTGGCAAAATAATAATTAGACAACAACTAAGTAGCCTAAGTAAATATGCTGAAGA 182
 Qy 392 tatatttggagaattattcaatgaagcacatagtttttcccttcagagtcacactcattgca 451
 Db 183 TATATTTGGAGAAATTATTCATTAAGCACATAGTATTTTCCCTTCAGAGTCAACTATTGCA 242
 Qy 452 agaactgtggacgcttcttctgttagtggttacacagcttgcacaaagggaagaatt 511
 Db 243 ACAACGTGTGGACCGTTTATCTGTGTAGTGTACACAGCTTGATCCAAAGGAGGAGAAAT 302
 Qy 512 gtctttgcaagataataacatagaggaagcgtttccggaagttctcaaatccaagaccagca 571
 Db 303 GTCCTTTGCAAGATATAACAATGAGGAAAGCTTTCGGAAGTTCTACAATTCAAGACCAGCA 362
 Qy 572 gcttttgcagacagacttgccttattccattacagagagcagtcagatgttttgaaca 631
 Db 363 GCTTTTCGATCGAAGACTTTTGCCTATTTCCTATTACAGAGAGCTACGATGTTTGTGAACA 422
 Qy 632 gctccactctcaataactactcactccttatagagatgattgtaagaaggtctgaagt 691
 Db 423 GCCCTCCACCTCTCAATATACCTACTCTCTTATAGAGATGATGCTAAAGAGGCTCTGAAGTT 482
 Qy 692 ttataccaactcttcgtatttcttcttgcattatgatactatggaa-agaataaattgtgcaagatacac 750
 Db 483 TTATACCAACTCTTCGTATTCTTTTGTATCTATGGAACAGACAACACTGTGTGAAGATACAG 542
 Qy 751 a 751
 Db 543 A 543

RESULT 15
 AV664100/c 547 bp mRNA EST 25-AUG-2000
 LOCUS AV664100 Bos taurus brain fetus Bos taurus cDNA clone EIBR040D11
 DEFINITION AV664100 5', mRNA sequence.
 AV664100
 AV664100.1 GI:9923130
 VERSION
 EST.
 KEYWORDS
 SOURCE COW.

ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 547)
 Sugimoto, Y., Hirotsune, S., Takasuga, A., Itoh, R., Jitohzono, A. and
 Suzuki, H.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shirakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugi@cocoa.ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.

FEATURES
 Location/Qualifiers
 1. 547

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1BR040D11"
/clone_lib="Bos taurus brain fetus"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/Note="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A
      was deleted from a NotI site"
BASE COUNT      118 a      95 c      198 g      134 t      2 others
ORIGIN
```

```

Query Match      18.8%; Score 493.6; DB 32; Length 547;
Best Local Similarity 96.0%; Pred. No. 2.5e-105;
Matches 505; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1300 cccccccacccacccagcagctgttgcagctccagcagcagcaccacccacg 1359
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
547 CACCCCCACCTCCACCTCCACCAANTGCTTTGCAAGCTCCAGCAGTACCACCTCCAG 488
QY 1360 ctctcttcagatgcccctggagttcttcacccagctctctccaaattgcacctctc 1419
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
487 CTCTCTTCAGATTGCCCTGGAGTTCTTCACCCAGCTCTCTCCCAATTGCACCTCTC 428
QY 1420 tagtacagcccttcacccagtagtagagctgccccagtagtgagactgtaccagttc 1479
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
427 TAGTACAGCCCTCTCCACCAAGTAGTAGAGCTGCCCCAGTATGTGAGACTGTACCAGTTC 368
QY 1480 atccactccacaaagtgaagttcaggggtgctcctccaccccccacccgctcctctgc 1539
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
367 ATCCACTCCACAAAGAGAAGTCCAGGGGCTGCCTCCACCCACCACCTCCCTTTGC 308
QY 1540 ctccacctggcattcgaccatcatcactgtcacagctgtgtcctcctcctccct 1599
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
307 CTCCACCTGGCATTAGACCATCATCACCCTGTACAGTTGCAGCTCTTGTCTATCCTCCCT 248
QY 1600 ctgggtacatccaactccaactctactgccccaggtccccaatgttccattaatgcctccat 1659
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
247 CTGGGTACATCCAACCTCCATCTACTGCCCCAGGTGCCCATGTTCATTAATGCTCCAT 188
QY 1660 ctctccatcaagttatcctgtctgtgagcaagccatccatcaacctaccctg 1719
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
187 CTCTCCATCACAAGTTACACCTGTCTTGTGAGCCAAAGCGTCATCCATCAACCCCTACCG 128
QY 1720 taatcagtgatgccaggagtggtgctactgaagcaatacgaagaggtattcagctacgca 1779
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
127 TATCAGTGACGCCAGAGGTACTTCTGGAGCATACGAAAGGTATTTCAGCTACGAA 68
QY 1780 aagtagaagagcagcgtagaacaggaagctaaagcatgaagcattga 1825
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
67 AAGTCGAAGAGACGCGTGAACAGGAAGCTAAACATGAACGCATTAA 22
```

1-10-1948

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:11:21 ; Search time 37.01 Seconds
(without alignments)
1998.339 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 559
Sequence: 1 MPLVKRNDPRHLCHTALPR.....AVEYSDSEDDSEFDVDMLE 559

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : SPTREMBL_16.*
- 1: sp-archaea.*
 - 2: sp-bacteria.*
 - 3: sp-fungi.*
 - 4: sp-human.*
 - 5: sp-invertebrate.*
 - 6: sp-mammal.*
 - 7: sp-mnc.*
 - 8: sp-organelle.*
 - 9: sp-phage.*
 - 10: sp-plant.*
 - 11: sp-rodent.*
 - 12: sp-unclassified.*
 - 13: sp-vertebrate.*
 - 14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	218	39.0	559	11 Q9ERQ9	Q9erq9 mus musculus
2	24	4.3	455	4 Q94974	Q94974 homo sapien
3	24	4.3	502	4 Q9UPY6	Q9upy6 homo sapien
4	22	3.9	496	4 Q9UDY7	Q9udy7 homo sapien
5	22	3.9	497	4 Q60794	Q60794 homo sapien
6	22	3.9	498	4 Q9Y6W5	Q9y6w5 homo sapien
7	12	2.1	210	4 Q9NKE9	Q9nke9 homo sapien
8	12	2.1	239	5 Q9W2R5	Q9w2r5 drosophila
9	12	2.1	280	6 Q9MYL6	Q9myl6 macaca neme
10	12	2.1	321	13 Q9PTD5	Q9ptd5 gallus gall
11	12	2.1	360	4 Q9NUZ7	Q9nuz7 homo sapien
12	12	2.1	409	10 Q9SBM1	Q9sbm1 volvox cart
13	12	2.1	446	4 Q9HCH4	Q9hch4 homo sapien
14	12	2.1	555	14 Q9YMX1	Q9ymx1 lymantria d
15	12	2.1	605	14 Q9Q5L3	Q9q5l3 cercopithec
16	12	2.1	802	11 P70433	P70433 mus musculus
17	12	2.1	1021	5 Q9XUV2	Q9xuv2 caenorhabdi
18	12	2.1	1527	11 Q9ES67	Q9es67 rattus norv
19	12	2.1	1822	4 Q9PLV7	Q9plv7 homo sapien

20	12	2.1	3394	4 Q9Y6V0	Q9y6v0 homo sapien
21	12	2.1	5120	13 Q9PU36	Q9pu36 gallus gall
22	11	2.0	61	11 Q08549	Q08549 mus musculus
23	11	2.0	268	13 Q0DGN3	Q0dgn3 gallus gall
24	11	2.0	433	10 Q22459	Q22459 chlamydomon
25	11	2.0	451	11 Q62775	Q62775 rattus norv
26	11	2.0	453	4 Q95763	Q95763 homo sapien
27	11	2.0	485	11 Q9Z0G8	Q9z0g8 rattus norv
28	11	2.0	532	11 Q9QY28	Q9qy28 mus musculus
29	11	2.0	613	5 Q9VKM2	Q9vkm2 drosophila
30	11	2.0	662	11 Q63376	Q63376 rattus norv
31	11	2.0	671	4 Q95559	Q95559 homo sapien
32	11	2.0	837	11 Q63618	Q63618 rattus norv
33	11	2.0	854	4 Q9H0A2	Q9h0a2 homo sapien
34	11	2.0	871	11 Q9ET47	Q9et47 mus musculus
35	11	2.0	1024	11 Q9ESJ5	Q9esj5 mus musculus
36	11	2.0	1037	13 Q73808	Q73808 fugu rubrip
37	11	2.0	1130	4 Q9H1V5	Q9h1v5 homo sapien
38	11	2.0	1148	4 Q99590	Q99590 homo sapien
39	11	2.0	1386	11 Q9R0I7	Q9r0i7 mus musculus
40	11	2.0	1388	4 Q9UEF3	Q9uef3 homo sapien
41	11	2.0	1542	4 Q9Y6X0	Q9y6x0 homo sapien
42	11	2.0	4833	11 Q9QYX6	Q9qyx6 mus musculus
43	11	2.0	5038	11 Q9QYX7	Q9qyx7 mus musculus
44	10	1.8	126	14 Q11299	Q11299 molluscum c
45	10	1.8	139	10 Q9SCM2	Q9scm2 arabidopsis

ALIGNMENTS

RESULT 1

Q9ERQ9 PRELIMINARY; PRT; 559 AA.
ID Q9ERQ9
AC Q9ERQ9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE WAVE-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428428; PubMed=10970852;
RA Westphal R.S., Soderling S.H., Alto N.M., Langeberg L.K., Scott J.D.;
RT "Scar/WAVE-1, a Wiskott-Aldrich syndrome protein, assembles an actin-
associated multi-kinase scaffold.";
RL EMBO J. 19:4589-4600(2000).
DR EMBL; AF290877; AAC02214.1; -
SQ SEQUENCE 559 AA; 61508 MW; 8746910987D80D16 CRC64;

Query Match 39.0%; Score 218; DB 11; Length 559;
Best Local Similarity 100.0%; Pred. No. 6e-218;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPLVKRNDPRHLCHTALPRGKNEECVTNISLANIIRQLSSLSKYAEDIFGELFNEAH 60
DB 1 MPLVKRNDPRHLCHTALPRGKNEECVTNISLANIIRQLSSLSKYAEDIFGELFNEAH 60
QY 61. SFSFRVNSLOERVDRLSVSVTQDLPKEEELSLODITMRKAFRSSTIQDQQLFDRKTLPIP 120
DB 61 SFSFRVNSLOERVDRLSVSVTQDLPKEEELSLODITMRKAFRSSTIQDQQLFDRKTLPIP 120
QY 121 LQETYDVCEGPPPLNITPTPRDDGEGELKFKYTNPSYFFDLWKEMKLODTEKREKRKQK 180
DB 121 LQETYDVCEGPPPLNITPTPRDDGEGELKFKYTNPSYFFDLWKEMKLODTEKREKRKQK 180
QY 181 QKNLDRPHEPEKVPRAHDRRRERWQKLAQGPELAEDDA 218
DB 181 QKNLDRPHEPEKVPRAHDRRRERWQKLAQGPELAEDDA 218

```

RESULT 2
O94974 PRELIMINARY; PRT; 455 AA.
AC O94974;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE KIAA0900 PROTEIN (FRAGMENT).
CN KIAA0900.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98402540; PubMed=9732292;
RA Bear J.E., Rawls J.F., Saxe C.L. III.;
RT "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT defects in late Dictyostelium development.";
RL J. Cell Biol. 142:1325-1335(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Machesky L.M., Insall R.H.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020707; BAA74923.1; -
DR EMBL; AF134305; AAD33054.1; -
DR InterPro; IPR002965; -
DR InterPro; IPR003124; -
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 455 AA; 50126 MW; 6135C2160991E8BB CRC64;

Query Match 4.3%; Score 24; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 4.4e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 154 PSYFFDLWKEKMLQDTEDKRKEK 177
Db 107 PSYFFDLWKEKMLQDTEDKRKEK 130

RESULT 3
O9UPY6 PRELIMINARY; PRT; 502 AA.
AC O9UPY6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE WASP-FAMILY PROTEIN.
CN WAVE3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99310608; PubMed=10381382;

```

```

RA Suetsugu S., Miki H., Takenawa T.;
RT "Identification of two human WAVE/SCAR homologues as general actin
RT regulatory molecules which associate with the Arp2/3 complex.";
RL Biochem. Biophys. Res. Commun. 260:296-302(1999).
DR EMBL; AB026543; BAA81796.1; -
DR InterPro; IPR002965; -
DR InterPro; IPR003124; -
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 502 AA; 55411 MW; 140DD58309345F35 CRC64;

Query Match 4.3%; Score 24; DB 4; Length 502;
Best Local Similarity 100.0%; Pred. No. 4.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 154 PSYFFDLWKEKMLQDTEDKRKEK 177
Db 154 PSYFFDLWKEKMLQDTEDKRKEK 177

RESULT 4
O9UDY7 PRELIMINARY; PRT; 496 AA.
AC O9UDY7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE SCAR2 (FRAGMENT).
CN SCAR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98402540; PubMed=9732292;
RA Bear J.E., Rawls J.F., Saxe C.L. III.;
RT "SCAR, a WASP-related protein, isolated as a suppressor of receptor
RT defects in late Dictyostelium development.";
RL J. Cell Biol. 142:1325-1335(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Machesky L.M., Insall R.H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF134304; AAD33053.2; -
DR InterPro; IPR002965; -
DR InterPro; IPR003124; -
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 496 AA; 53963 MW; 8A9D97D0D964734F CRC64;

Query Match 3.9%; Score 22; DB 4; Length 496;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 529 NDVATILSRRIAYEYSDSDS 550
Db 465 NDVATILSRRIAYEYSDSDS 486

RESULT 5
O60794 PRELIMINARY; PRT; 497 AA.
AC O60794;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE DJ393P12.2 (HYPOTHETICAL PROLINE-RICH PROTEIN KIAA0269 LIKE)

```

DE (FRAGMENT).
GN DJ393P12.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP Bird C.;
RA Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RL EMBL; AL022578; CAA18609.1; -
DR InterPro: IPR003124; -
DR Pfam: PF02205; WH2; 1.
DR SMART: SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 497 AA; 54062 MW; 58872599FDF63A6B CRC64;

Query Match 3.9%; Score 22; DB 4; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.7e-14; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 529 NDVATILSRRIAYEYSDSEDS 550
Db 466 NDVATILSRRIAYEYSDSEDS 487
|||||

RESULT 6
QY6W5 PRELIMINARY; PRT; 498 AA.
AC QY6W5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE WASP-FAMILY PROTEIN (DJ144C9.2).
GN WAVE2 OR DJ144C9.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE-99310608; PubMed-10381382;
RA Suetsugu S., Miki H., Takenawa T.;
RT "Identification of two human WAVE/SCAR homologues as general actin
regulatory molecules which associate with the Arp2/3 complex.";
RL Biochem. Biophys. Res. Commun. 260:296-302(1999).
[2]
RN [1]
RP SEQUENCE FROM N.A.
RA Hail R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RX EMBL; AB026542; BAA81795.1; -
DR EMBL; AL096774; CAC18518.1; -
DR InterPro: IPR002965; -
DR InterPro: IPR002966; -
DR InterPro: IPR003124; -
DR Pfam: PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHXPENS.
DR PRINTS; PR01218; PSTLXTENSIN.
DR SMART: SM00246; WH2; 1.
SQ SEQUENCE 498 AA; 54283 MW; C737CE963016DE94 CRC64;

Query Match 3.9%; Score 22; DB 4; Length 498;
Best Local Similarity 100.0%; Pred. No. 5.7e-14; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 529 NDVATILSRRIAYEYSDSEDS 550
Db 467 NDVATILSRRIAYEYSDSEDS 488
|||||

RESULT 7

Q9NXE9
ID Q9NXE9 PRELIMINARY; PRT; 210 AA.
AC Q9NXE9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-NAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CDNA FLJ20289 FIS, CLONE HEP04492.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Odayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK000296; BAA91064.1; -
DR InterPro: IPR002965; -
DR PRINTS; PR01217; PRICHXPENS.
SQ SEQUENCE 210 AA; 22657 MW; FC1E09F8AE83C082 CRC64;

Query Match 2.1%; Score 12; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 0.00068; Indels 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 0;

QY 424 PPPPPPPPLPPP 435
Db 35 PPPPPPPPLPPP 46
|||||

RESULT 8
Q9W2R5 PRELIMINARY; PRT; 239 AA.
ID Q9W2R5;
AC Q9W2R5;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-WAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG15225 PROTEIN.
GN CG15225.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP STRALINE-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Chertis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003452; AAF46625.1; -
 DR FlyBase; FBgn0034551; CG15225.
 DR InterPro; IPR002965; -
 DR PRINTS; PR01217; PRICHEXTENS.
 SQ SEQUENCE 239 AA; 26175 MW; 81EEEB5DD2FC5FB7 CRC64;

Query Match 2.1%; Score 12; DB 5; Length 239;
 Best Local Similarity 100.0%; Pred. No. 0.00077;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 424 PPPPPPPPLPPP 435

DB 202 PPPPPPPPLPPP 213

RESULT 9

O9MYL6

ID O9MYL6 PRELIMINARY; PRT; 280 AA.

AC O9MYL6;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE FAS LIGAND.

CN PT-FASL OR CW-FASL OR RW-FASL.

OS Macaca nemestrina (Pig-tailed macaque),

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), and

OS Macaca mulatta (Rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;

OC Cercopitheciinae; Macaca.

OX NCBI_TaxID=9545, 9541, 9544;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=M.nemestrina; STRAIN=PIG-TAILED MONKEY;

RA Kirii Y., Inoue T., Yoshino K.;

RT "Pig-tailed monkey Fas ligand mRNA, complete cds.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=M.fascicularis; STRAIN=CYNOMOLGUS MONKEY;

RA Kirii Y., Inoue T., Yoshino K.;

RT "Cynomolgus monkey Fas ligand mRNA, complete cds.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=M.mulatta; STRAIN=RHESUS MONKEY;

RA Kirii Y., Inoue T., Yoshino K.;

RT "Rhesus monkey Fas ligand mRNA, complete cds.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB035140; BAA90296.1; -

DR EMBL; AB035138; BAA90294.1; -

DR EMBL; AB035139; BAA90295.1; -

DR InterPro; IPR000478; -

DR pfam; PF00229; TNF_1.

DR PRINTS; PR01234; TNCRSISFCT.

DR PROSITE; PS00251; TNF_1; 1.

DR PROSITE; PS00049; TNF_2; 1.
 SQ SEQUENCE 280 AA; 31367 MW; FOB284D61A132EB4 CRC64;

Query Match 2.1%; Score 12; DB 6; Length 280;

Best Local Similarity 100.0%; Pred. No. 0.00088;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 424 PPPPPPPPLPPP 435

DB 46 PPPPPPPPLPPP 57

RESULT 10

O9PTD5

ID O9PTD5 PRELIMINARY; PRT; 321 AA.

AC O9PTD5;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE KIAA0009 PROTEIN.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RA Castagnola P., Tonachini L., Monticone M.;

RT "cDNA cloning of chick KIAA0009.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF208489; AAF21014.1; -

DR InterPro; IPR001799; -

DR ProDom; PD002533; -

SQ SEQUENCE 321 AA; 35667 MW; 624AC2AC9AC099B0 CRC64;

Query Match 2.1%; Score 12; DB 13; Length 321;

Best Local Similarity 100.0%; Pred. No. 0.001;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 424 PPPPPPPPLPPP 435

DB 193 PPPPPPPPLPPP 204

RESULT 11

O9NUZ7

ID O9NUZ7 PRELIMINARY; PRT; 360 AA.

AC O9NUZ7;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE CDNA FLJ11029 FIS, CLONE PLACE1004156.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,

RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,

RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK001891; BAA91964.1; -

DR InterPro; IPR002965; -

DR PRINTS; PR01217; PRICHEXTENS.

SQ SEQUENCE 360 AA; 40154 MW; 55747432CA0C53DA CRC64;


```

RL DNA Res. 7:273-281(2000).
EMBL AB046818; BAB13424.1; -.
FT NON_TER 1
SQ SEQUENCE 446 AA; 51473 MW; 5C82402E11E75570 CRC64;

Query Match 2.1%; Score 12; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 P P P P P P P P P P P P 435
DB 344 P P P P P P P P P P P P 355

RESULT 14
Q9YMX1 PRELIMINARY; PRT; 555 AA.
ID AC Q9YMX1
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2001 (TrEMBLrel. 16, Last annotation update)
DE ESSENTIAL STRUCTURAL PROTEIN PP78-81.
OS Lymantria dispar multicapsid nuclear polyhedrosis virus (LGMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OC NCBI_TaxID=10449;
OX RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99124785; PubMed=9887315;
RA Kuzio J., Pearson M.N., Harwood S.H., Evans J.T.,
RA Slavicek J.M., Rohrmann G.F.;
RT "Sequence and analysis of the genome of a baculovirus pathogenic for
RT Lymantria dispar.";
RL Virology 253:17-34(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX Kuzio J., Pearson M.N., Harwood S.H., Evans J.T.,
RA Slavicek J., Rohrmann G.F.;
RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX Kuzio J.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF081810; AAC70187.1; -.
DR InterPro: IPR001005; -.
DR InterPro: IPR002965; -.
DR InterPro: IPR003124; -.
DR Pfam: PF02205; WH2; 3.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS000037; MYB_1; UNKNOWN_1.
DR SMART: SM00246; WH2; 1.
SQ SEQUENCE 555 AA; 61055 MW; 6B638D6E02279887 CRC64;

Query Match 2.1%; Score 12; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 P P P P P P P P P P P P 436
DB 268 P P P P P P P P P P P P 279

RESULT 15
Q9Q5L3 PRELIMINARY; PRT; 605 AA.
ID AC Q9Q5L3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE EBNA2-LIKE PROTEIN.
OS cercopithicine herpesvirus 15.

```

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=104228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057932; Pubmed=10590127;
 RA Peng R., Gordanze A.V., Fuentes Panana E.M., Wang F., Zong J.,
 RA Hayward G.S., Tan J., Ling P.D.;
 RT "Sequence and functional analysis of EBNA-LP and EBNA2 proteins from
 RT nonhuman primate lymphocryptoviruses.";
 RL J. Virol. 74:379-389(2000).
 DR EMBL; AF200187; AAF25206.1; -.
 DR InterPro; IPR000633; -.
 DR InterPro; IPR002965; -.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR PRINTS; PR00806; VINCULIN.
 SQ SEQUENCE 605 AA; 64712 MW; AD0F77111208AA4E CRC64;

Query Match 2.1%; Score 12; DB 14; Length 605;
 Best Local Similarity 100.0%; Pred.No. 0.0017;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 424 P P P P P P P P L P P P 435
 Db 66 P P P P P P P P L P P P 77

Search completed: June 20, 2001, 12:14:38
 Job time: 197 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:10:01 ; Search time 22.54 Seconds
(without alignments)
1889.156 Million cell updates/sec

Title: US-09-425-501-2

Perfect score: 559
Sequence: 1 MPLVKNIDPRHLCHLTALPR.....AVEYSDSEDDSEFDEVDWLE 559

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	2.1	281	I38707	Fas ligand - human
2	12	2.1	389	D27200	proline-rich prote
3	12	2.1	555	T30349	structural protein
4	12	2.1	1021	T23252	hypothetical prote
5	11	2.0	433	T07910	hydroxyproline-ric
6	11	2.0	477	I38409	adenyl cyclase-a
7	11	2.0	662	D40228	neurexin II-beta p
8	11	2.0	798	D96563	probable bZIP prot
9	11	2.0	994	S19595	chloride channel p
10	11	2.0	1148	T09073	splicing factor Si
11	11	2.0	1388	T00063	hypothetical prote
12	10	1.8	139	T46186	hypothetical prote
13	10	1.8	145	T48552	glutaredoxin-like
14	10	1.8	172	D41132	collagen-related p
15	10	1.8	182	T30760	hypothetical prote
16	10	1.8	185	B26659	nodulin-20 precurs
17	10	1.8	207	I53154	scleraxis - mouse
18	10	1.8	211	S28304	hypothetical prote
19	10	1.8	275	T02559	probable spliceoso
20	10	1.8	367	S02193	cellular tumor ant
21	10	1.8	442	T34018	hypothetical prote
22	10	1.8	457	S41318	hypothetical prote
23	10	1.8	473	B85187	glycoprotein homol
24	10	1.8	477	JC4386	adenyl cyclase-a
25	10	1.8	551	A60047	serine/arginine-ri
26	10	1.8	681	JC5929	hypothetical prote
27	10	1.8	691	T46476	hypothetical prote
28	10	1.8	907	E96636	hypothetical prote
29	10	1.8	1058	T13286	cappuccino gene pr

30	10	1.8	1064	2	T13963	formin related pro
31	10	1.8	1108	2	A48508	cyclic-nucleotide
32	10	1.8	1171	2	T17454	diaphanous-related
33	10	1.8	1201	2	G86441	unknown protein fi
34	10	1.8	1206	2	S24407	formin isoform IV
35	10	1.8	1255	2	T31065	diaphanous protein
36	10	1.8	1375	2	S48375	hypothetical prote
37	10	1.8	1468	2	S11515	formin - mouse
38	10	1.8	1872	2	T24683	hypothetical prote
39	10	1.8	2783	1	A41948	alpha-fetoprotein
40	9	1.6	100	2	T17126	hypothetical prote
41	9	1.6	135	2	E96750	hypothetical prote
42	9	1.6	154	2	T41831	AcMNPV orf91 - Bom
43	9	1.6	161	2	H85120	probable proline-r
44	9	1.6	194	2	A38203	proline-rich prote
45	9	1.6	195	2	T07735	nodulin-20a - soyb

ALIGNMENTS

RESULT 1
I38707
Fas ligand - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38707; JC2340; S57565; I38554
R:Takahashi, T.; Tanaka, M.; Inazawa, J.; Abe, T.; Suda, T.; Nagata, S.
Int. Immunol. 6, 1567-1574, 1994
A:Title: Human Fas ligand: gene structure, chromosomal location and species specific
A:Reference number: I38707; MUID:95127560
A:Accession: I38707
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-281 <RES>
A:Cross-references: EMBL:U11821; NID:G9595430; PIDN:AAC50124.1.; PID:G9595431
R:Mita, E.; Hayashi, N.; Iio, S.; Takehara, T.; Hijioka, T.; Kasahara, A.; Fusamoto, Biochem. Biophys. Res. Commun. 204, 468-474, 1994
A:Title: Role of Fas ligand in apoptosis induced by hepatitis C virus infection.
A:Reference number: JC2340; MUID:95071350
A:Accession: JC2340
A:Molecule type: DNA
A:Residues: 1-281 <MIT>
A:Cross-references: GB:D38122; DBDJ:D29820; NID:G601892; PIDN:BAA07320.1.; PID:G136990
R:Schatzlein, C.E.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57565
A:Accession: S57565
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-281 <SCH>
A:Cross-references: EMBL:X89102; NID:G887455; PID:G887456
R:Alderson, M.R.; Tough, T.W.; Davis-Smith, T.; Braddy, S.; Falk, B.; Schooley, K.A.; J. Exp. Med. 181, 71-77, 1995
A:Title: Fas ligand mediates activation-induced cell death in human T lymphocytes.
A:Reference number: I38554; MUID:95105731
A:Accession: I38554
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-281 <RE2>
A:Cross-references: EMBL:U08137; NID:G624627; PIDN:AAC50071.1.; PID:G624628
C:Genetics:
A:Gene: FasL
A:Introns: 151/1; 116/3
C:Keywords: glycoprotein; transmembrane protein
F:80-102/Domains: transmembrane #status predicted <TMM>
F:76,184,250,260/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.1%; Score 12; DB 2; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
 Db 46 PPPPPPPPLPPP 57

RESULT 2

S27200
 proline-rich protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
 C:Accession: S27200
 R:Sazuka, T.; Tomooka, Y.; Kathju, S.; Ikawa, Y.; Noda, M.; Kumar, S.
 Biochim. Biophys. Acta 1132, 240-248, 1992
 A:Title: Identification of a developmentally regulated gene in the mouse central nervous system
 A:Reference number: S27200; MUID:93041923
 A:Accession: S27200
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-389 <SA2>
 A:Cross-references: GB:D10727; NID:g220499; PIDN:BAA01570.1; PID:g1002045; PID:g220500

Query Match 2.1%; Score 12; DB 2; Length 389;
 Best Local Similarity 100.0%; Pred. No. 0.0019;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
 Db 37 PPPPPPPPLPPP 48

RESULT 3

T30349
 structural protein pp78-81 - Lymantria dispar nuclear polyhedrosis virus
 C:Species: Lymantria dispar nuclear polyhedrosis virus, IdMNPV
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T30349
 R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohrer, J.
 Virology 253, 17-34, 1999
 A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar
 A:Reference number: 220836; MUID:99124785
 A:Accession: T30349
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-555 <KU2>
 A:Cross-references: EMBL:AF081810; PIDN:AAC70187.1

Query Match 2.1%; Score 12; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 0.0026;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 PPPPPPPPLPPP 436
 Db 268 PPPPPPPPLPPP 279

RESULT 4

T23252
 hypothetical protein K02E2.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T23252
 R:Lloyd, C.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19716
 A:Accession: T23252
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-1021 <WIL>
 A:Cross-references: EMBL:Z81560; PIDN:CAB04547.1; GSPDB:GN00023; CESP:K02E2.2
 A:Experimental source: clone K02E2
 C:Genetics:

A:Gene: CESP:K02E2.2
 A:Map position: 5
 A:Introns: 41/1; 79/3; 131/3; 197/1; 262/1; 297/3; 566/3; 610/3; 735/3; 775/3; 834/3;

Query Match 2.1%; Score 12; DB 2; Length 1021;
 Best Local Similarity 100.0%; Pred. No. 0.0047;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 PPPPPPPPLPPP 359
 Db 787 PPPPPPPPLPPP 798

RESULT 5

T07910
 hydroxyproline-rich glycoprotein GAS29 precursor - Chlamydomonas reinhardtii
 C:Species: Chlamydomonas reinhardtii
 C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999
 C:Accession: T07910
 R:Rodriguez-Martinez, H.; Haring, M.A.; Von Gromof, E.; Beck, C.F.
 submitted to the EMBL Data Library, July 1997
 A:Reference number: Z16207
 A:Accession: T07910
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-433 <ROD>
 A:Cross-references: EMBL:AF015884; NID:g2384729; PIDN:AAB69863.1; PID:g2384730
 A:Experimental source: gametes
 C:Genetics:

A:Gene: GAS29
 F:1-31/Domain: signal sequence #status predicted <SIG>
 F:32-433/Product: hydroxyproline-rich glycoprotein GAS29 #status predicted <MAP>

Query Match 2.0%; Score 11; DB 2; Length 433;
 Best Local Similarity 100.0%; Pred. No. 0.02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 434
 Db 63 PPPPPPPPLPPP 73

RESULT 6

I38409
 adenyl cyclase-associated protein CAP2 - human
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 19-May-2000
 C:Accession: I38409
 R:Yu, G.; Swiston, J.; Young, D.
 J. Cell Sci. 107, 1671-1678, 1994
 A:Title: Comparison of human CAP and CAP2, homologs of the yeast adenyl cyclase-adenyl cyclase
 A:Reference number: I38409; MUID:95051124
 A:Accession: I38409
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-477 <RES>
 A:Cross-references: EMBL:U02390; NID:g409928; PIDN:AAA20587.1; PID:g409929
 C:Genetics:
 A:Gene: GDB:CAP2
 A:Cross-references: GDB:9957118
 C:Superfamily: adenyl cyclase-associated protein MCH1

Query Match 2.0%; Score 11; DB 2; Length 477;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PPPPPPPPLPPP 436
 Db 232 PPPPPPPPLPPP 242

RESULT 7

D40228
 neurexin II-beta precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
 C:Accession: D40228; S27888
 R:Ushakov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
 Science 257, 50-56, 1992
 A:Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor
 A:Reference number: A40228; MUID:92320296
 A:Accession: D40228
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-662 <XST>
 A:Cross-references: GB:M96377; NID:q205717; PIDN:AAA1708.1; PID:q205719
 A:Note: authors translated GAC for residue 411 as Thr and ACC for residue 412 as Asp
 C:Keywords: alternative splicing; transmembrane protein
 F:1-47/Domain: signal sequence; status predicted <SIG>
 F:48-662/Product: neurexin II-beta #status predicted <MAT>

Query Match 2.0%; Score 11; DB 2; Length 662;

Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 LPPPPPPPLP 433

|||||

DB 22 LPPPPPPPLP 32

RESULT 8

D96563
 probable bZIP protein, 48652-45869 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96563
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 Jensen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D96563
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-798 <STO>
 A:Cross-references: GB:AE005173; NID:g10645444; PIDN:AAG21560.1; GSPDB:GNO0141
 C:Genetics:
 A:Gene: FL9K6.7
 A:Map position: 1

Query Match 2.0%; Score 11; DB 2; Length 798;

Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 LPPPPPPPLP 433

|||||

DB 115 LPPPPPPPLP 125

RESULT 9

S19595
 chloride channel protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
 C:Accession: S19595

R.Steinmeyer, K.; Ortland, C.; Jentsch, T.J.
 Nature 354, 301-304, 1991
 A:Title: Primary structure and functional expression of a developmentally regulated s
 A:Reference number: S19595; MUID:92065954
 A:Accession: S19595
 A:Molecule type: mRNA
 A:Residues: 1-994 <ST>
 A:Cross-references: GB:X62894; NID:g57744; PIDN:CAA44683.1; PID:g57745
 C:Superfamily: CBS homology
 C:Keywords: transmembrane protein
 F:830-877/Domain: CBS homology <CBS>

Query Match 2.0%; Score 11; DB 2; Length 994;

Best Local Similarity 100.0%; Pred. No. 0.043;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLP 434

|||||

DB 728 PPPPPPPPLP 738

RESULT 10

T09073
 splicing factor Sipl - human
 C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: T09073
 R:Zhang, W.J.; Wu, J.Y.
 Mol. Cell. Biol. 18, 676-684, 1998
 A:Title: Sipl, a novel RS domain-containing protein essential for pre-mRNA splicing.
 A:Reference number: Z16554; MUID:98107652
 A:Accession: T09073
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1148 <ZHA>
 A:Cross-references: EMBL:AF030234; NID:g2822459; PIDN:AAC39565.1; PID:g2822460
 A:Experimental source: HeLa cells
 C:Keywords: pre-mRNA splicing

Query Match 2.0%; Score 11; DB 2; Length 1148;

Best Local Similarity 100.0%; Pred. No. 0.05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 QGLPPPPPPPP 431

|||||

DB 959 QGLPPPPPPPP 969

RESULT 11

T00063
 hypothetical protein KIAA0437 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
 C:Accession: T00063
 R:Shikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka,
 DNA Res. 4, 307-313, 1997
 A:Title: Prediction of the coding sequences of unidentified human genes. VIII. 78 new
 A:Reference number: Z14084; MUID:98116655
 A:Accession: T00063
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1388 <ISH>
 A:Cross-references: EMBL:AB007897; NID:g2879924; PIDN:BAA24826.1; PID:g2879925
 A:Experimental source: brain
 C:Genetics:
 A:Note: KIAA0437

Query Match 2.0%; Score 11; DB 2; Length 1388;

Best Local Similarity 100.0%; Pred. No. 0.059;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 PPPPPPLPPP 435
Db 1315 PPPPPPLPPP 1325

RESULT 12

T46186
hypothetical protein T8H10.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46186
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223014
A:Accession: T46186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-139 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Introns: 20/1
A:Note: T8H10.100

Query Match 1.8%; Score 10; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 PPPPPPPATA 362
Db 116 PPPPPPPATA 125

RESULT 13

T48552
glutaredoxin-like protein - Arabidopsis thaliana
N:Alternate names: protein F14F18.100
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48552
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224490
A:Accession: T48552
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <BEV>
A:Cross-references: EMBL:AL163812
A:Experimental source: cultivar Columbia; BAC clone F14F18
C:Genetics:
A:Map position: 5
A:Note: F14F18.100

Query Match 1.8%; Score 10; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PPPPPPLPPP 435
Db 20 PPPPPPLPPP 29

RESULT 14

D41132
collagen-related protein 4 - Hydra magnipapillata (fragment)
C:Species: Hydra magnipapillata
C:Date: 05-Jun-1992 #sequence_revision 12-Jun-1992 #text_change 01-Dec-2000
C:Accession: D41132; S21932
R:Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.

J. Cell Biol. 115, 1159-1169, 1991
A:Title: Mini-collagens in hydra nematocytes.
A:Reference number: A41132; MUID:92064646
A:Accession: D41132
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-172 <KUR>
A:Cross-references: EMBL:X61048; NID:g9452; PIDN:CAA43382.1; PID:g9453
A:Note: the authors translated the codon GGT for residue 142 as Pro, and TTA for resi
A:Note: submitted to the EMBL data Library, July 1991
C:Superfamily: unassigned collagens

Query Match 1.8%; Score 10; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 GLPPPPPPPP 431
Db 142 GLPPPPPPPP 151

RESULT 15

T30760
hypothetical protein 158R - Molluscum contagiosum virus 1
N:Alternate names: MC158R
C:Species: Molluscum contagiosum virus 1
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-May-2000
C:Accession: T30760
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host
A:Reference number: 220876; MUID:96325459
A:Accession: T30760
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-182 <SEN>
A:Cross-references: EMBL:U60315; PIDN:AAC55286.1
C:Genetics:
A:Note: MC158R

Query Match 1.8%; Score 10; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 PPPPPPLPLP 433
Db 104 PPPPPPLPLP 113

Search completed: June 20, 2001, 12:12:03
Job time: 122 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 12:09:11 ; Search time 13.93 Seconds
(without alignments)
808.388 Million cell updates/sec

Title: US-09-425-501-2
Perfect score: 559
Sequence: 1 MPLVKNIDPHRLCHTALPR.....AVEYSDSEDDSEFDEVDWLE 559

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 193259 seqs, 20144635 residues

Word size : 0

Total number of hits satisfying Chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	2.1	281	2	US-08-810-453-2
2	12	2.1	281	3	US-08-815-190A-2
3	12	2.1	281	4	US-09-290-640-25
4	12	2.1	281	5	PCT-US95-00362-2
5	12	2.1	311	4	US-09-179-558-66
6	11	2.0	18	4	US-08-602-999A-320
7	11	2.0	28	2	US-08-769-745-39
8	10	1.8	18	4	US-08-602-999A-409
9	10	1.8	20	4	US-08-602-999A-358
10	10	1.8	288	3	US-08-545-196B-21
11	10	1.8	288	4	US-09-028-327-3
12	10	1.8	480	3	US-09-189-035-5
13	10	1.8	480	4	US-09-382-086-5
14	10	1.8	539	3	US-08-906-360-1
15	10	1.8	688	2	US-09-016-000-2
16	10	1.8	1248	2	US-09-080-897-2
17	10	1.8	1248	4	US-09-323-735-2
18	10	1.8	1255	2	US-09-080-897-4
19	10	1.8	1255	4	US-08-899-595-1
20	10	1.8	1255	4	US-09-323-735-4
21	10	1.8	1315	4	US-08-899-595-3
22	9	1.6	17	4	US-08-602-999A-360
23	9	1.6	17	4	US-08-602-999A-374
24	9	1.6	23	1	US-08-268-251-35
25	9	1.6	23	5	PCT-US93-01112-35
26	9	1.6	338	1	US-08-218-686-2
27	9	1.6	338	3	US-08-460-242-2

28	9	1.6	393	2	US-09-026-587-3	Sequence 3, Appli
29	9	1.6	393	2	US-09-227-420-3	Sequence 3, Appli
30	9	1.6	450	3	US-09-120-365-68	Sequence 68, Appl
31	9	1.6	450	4	US-09-515-039-68	Sequence 68, Appl
32	9	1.6	487	1	US-08-452-722-7	Sequence 7, Appli
33	9	1.6	487	1	US-08-404-731A-7	Sequence 7, Appli
34	9	1.6	487	1	US-08-344-227-7	Sequence 7, Appli
35	9	1.6	487	2	US-08-503-226B-7	Sequence 7, Appli
36	9	1.6	487	3	US-08-721-458B-7	Sequence 7, Appli
37	9	1.6	905	2	US-08-574-959A-9	Sequence 9, Appli
38	9	1.6	1135	2	US-08-574-959A-7	Sequence 7, Appli
39	9	1.6	1612	3	US-08-545-860D-48	Sequence 48, Appl
40	9	1.6	1612	5	PCT-US94-04496-48	Sequence 48, Appl
41	9	1.6	1829	4	US-09-157-420-1	Sequence 1, Appli
42	9	1.6	1865	1	US-08-588-985-2	Sequence 2, Appli
43	9	1.6	1865	1	US-08-971-988-2	Sequence 2, Appli
44	9	1.6	1958	1	US-07-945-283-2	Sequence 2, Appli
45	9	1.6	3118	2	US-08-457-273B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-810-453-2
; Sequence 2, Application US/08810453
; Patent No. 5858990
; GENERAL INFORMATION:
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: FAS LIGAND COMPOSITIONS FOR TREATMENT OF
; TITLE OF INVENTION: PROLIFERATIVE DISORDERS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: Federal Reserve Plaza, 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/810,453
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: S1237/7004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; TELEX: 343
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 281 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-810-453-2

Query Match 2.1%, Score 12; DB 2; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 424 P P P P P P P P P P P P P P 435
Db 46 P P P P P P P P P P P P P P 57

RESULT 2
US-08-815-190A-2
; Sequence 2, Application US/08815190A
; Patent No. 6046310
; GENERAL INFORMATION:
; APPLICANT: Queen, Cary L.
; APPLICANT: Schneider, William P.
; APPLICANT: Vasquez, Maximiliano
; TITLE OF INVENTION: Fas Ligand Fusion Proteins and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/815,190A
; FILING DATE: 11-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/614,584
; FILING DATE: 13-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 011823-0067100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 281 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-815-190A-2

Query Match 2.1%; Score 12; DB 3; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
DB 46 PPPPPPPPLPPP 57

RESULT 3
US-09-290-640-25
; Sequence 25, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcusson, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; CURRENT FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-290-640-25

Query Match 2.1%; Score 12; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
DB 46 PPPPPPPPLPPP 57

RESULT 4
PCT-US95-00362-2
; Sequence 2, Application PC/TUS9500362
; GENERAL INFORMATION:
; APPLICANT: IMMUNEX CORPORATION
; TITLE OF INVENTION: Ligand That Binds Fas Antigen
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00362
; FILING DATE: 06-JAN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,138
; FILING DATE: 07-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/190,559
; FILING DATE: 01-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Anderson, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2805-WO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 281 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-00362-2

Query Match 2.1%; Score 12; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
DB 46 PPPPPPPPLPPP 57

RESULT 5
US-09-179-558-66
; Sequence 66, Application US/09179558
; Patent No. 6180612
; GENERAL INFORMATION:
; APPLICANT: Hockensmith, Joel W.
; APPLICANT: Muthuswami, Rohini
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TARGETING DNA METABOLIC PROCESSES USING
; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES

NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/179,558
FILING DATE: 27-OCT-1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 09/060,470
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 60/063,898
FILING DATE: 31-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 9426-005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-179-558-66

Query Match 2.1%; Score 12; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 435
DB 8 PPPPPPPPLPPP 19
|||||

RESULT 6
US-08-602-999A-320
Sequence 320, Application US/08602999A
Patent No. 6184205
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: KAY, Brian K.
APPLICANT: THORN, Judith M.
APPLICANT: GUILLIAM, Lawrence A.
APPLICANT: DER, Channing J.
APPLICANT: FOWLKES, Dana M.
APPLICANT: RIDER, James E.
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
TITLE OF INVENTION: ISOLATING AND USING SAME
NUMBER OF SEQUENCES: 467
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,999A
FILING DATE: 16-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-202
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 320:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-602-999A-320

Query Match 2.0%; Score 11; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 434
DB 5 PPPPPPPPLPPP 15
|||||

RESULT 7
US-08-769-745-39
Sequence 39, Application US/08769745
Patent No. 5955259
GENERAL INFORMATION:
APPLICANT: Holmes, Todd C.
APPLICANT: Levitan, Irwin B.
APPLICANT: Brandeis University
TITLE OF INVENTION: Mechanism for the Regulation of Ion
TITLE OF INVENTION: Channel Activity
FILE REFERENCE: BRU96-02
CURRENT APPLICATION NUMBER: US/08/769,745
CURRENT FILING DATE: 1996-12-19
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 39
LENGTH: 28
TYPE: PRT
ORGANISM: Rat
US-08-769-745-39

Query Match 2.0%; Score 11; DB 2; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLPPP 434
DB 6 PPPPPPPPLPPP 16
|||||

RESULT 8
US-08-602-999A-409
Sequence 409, Application US/08602999A
Patent No. 6184205
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: KAY, Brian K.
APPLICANT: THORN, Judith M.

APPLICANT: QUILLIAM, Lawrence A.
APPLICANT: DER, Channing J.
APPLICANT: FOWLKES, Dana M.
APPLICANT: RIDER, James E.
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
TITLE OF INVENTION: ISOLATING AND USING SAME
NUMBER OF SEQUENCES: 467
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,999A
FILING DATE: 16-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-202
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
INFORMATION FOR SEQ ID NO: 409:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-602-999A-409

Query Match 1.8%; Score 10; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLP 433
| | | | | | | | | |
DB 6 PPPPPPPPLP 15

RESULT 9
US-08-602-999A-358
Sequence 358, Application US/08602999A
Patent No. 6184205
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: KAY, Brian K.
APPLICANT: THORN, Judith M.
APPLICANT: QUILLIAM, Lawrence A.
APPLICANT: DER, Channing J.
APPLICANT: FOWLKES, Dana M.
APPLICANT: RIDER, James E.
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
TITLE OF INVENTION: ISOLATING AND USING SAME.
NUMBER OF SEQUENCES: 467
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,999A
FILING DATE: 16-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-202
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
INFORMATION FOR SEQ ID NO: 358:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-602-999A-358

Query Match 1.8%; Score 10; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPPPLP 433
| | | | | | | | | |
DB 7 PPPPPPPPLP 16

RESULT 10
US-08-545-196B-21
Sequence 21, Application US/08545196B
Patent No. 6080577
GENERAL INFORMATION:
APPLICANT: MELKI, JUDITH
APPLICANT: MUNNICH, ARNOLD
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,196B
FILING DATE: 19-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FARACI, C. J.
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 2121-110P
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-545-196B-21

Query Match 1.8%; Score 10; DB 3; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PPPPPPLPPP 435
| | | | | | | | | |
Db 212 PPPPPPLPPP 221

RESULT 11
US-09-028-327-3
; Sequence 3, Application US/09028327
; Patent No. 6130064
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN SMN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,327
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0482 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1857114
US-09-028-327-3

Query Match 1.8%; Score 10; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PPPPPPLPPP 435
| | | | | | | | | |
Db 212 PPPPPPLPPP 221

RESULT 12
US-09-189-035-5
; Sequence 5, Application US/09189035
; Patent No. 6020165

; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
; FILE REFERENCE: PF-0638 US
; CURRENT APPLICATION NUMBER: US/09/189,035
; CURRENT FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: g2245671
US-09-189-035-5

Query Match 1.8%; Score 10; DB 3; Length 480;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 SPTPPPPPPP 330
| | | | | | | | | |
Db 364 SPTPPPPPPP 373

RESULT 13
US-09-382-086-5
; Sequence 5, Application US/09382086
; Patent No. 6201106
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
; FILE REFERENCE: PF-0638 US
; CURRENT APPLICATION NUMBER: US/09/382,086
; CURRENT FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 09/189,035
; EARLIER FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: g2245671
US-09-382-086-5

Query Match 1.8%; Score 10; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 SPTPPPPPPP 330
| | | | | | | | | |
Db 364 SPTPPPPPPP 373

RESULT 14
US-08-906-360-1
; Sequence 1, Application US/08906360
; Patent No. 6013441
; GENERAL INFORMATION:
; APPLICANT: Meruelo, Daniel
; APPLICANT: Pampeno, Christine
; TITLE OF INVENTION: MAMMALIAN HUMAN FXI-T1
; FILE REFERENCE: 8105-010

; CURRENT APPLICATION NUMBER: US/08/906,360
 ; EARLIER FILING DATE: 1997-08-05
 ; EARLIER APPLICATION NUMBER: 60/023,173
 ; EARLIER FILING DATE: 1996-08-05
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 539
 ; TYPE: PRT
 ; ORGANISM: mammalian
 US-08-906-360-1

Query Match 1.8%; Score 10; DB 3; Length 539;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 PPPPPPLPP 434
 Db 148 PPPPPPLPP 157

RESULT 15
 US-09-016-000-2
 ; Sequence 2, Application US/09016000
 ; Patent No. 5962232
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Akerblom, Ingrid E.
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Guegler, Karl G.
 ; TITLE OF INVENTION: PROTEIN KINASE MOLECULES
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/016,000
 ; FILING DATE: HEREWITH
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0465 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 688 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: HEARNOT01
 ; CLONE: 307624
 US-09-016-000-2

Query Match 1.8%; Score 10; DB 2; Length 688;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 PPPPPPLPP 433
 Db 29 PPPPPPLPP 38

Search completed: June 20, 2001, 12:11:35
 Job time: 144 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 18:20:12 ; Search time 3470.92 Seconds
(without alignments)
11698.006 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 ctctcttgcaacttgcggat.....aatggttcattttaaagtt 2625

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pri:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pri0:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2625	100.0	2625	91	D87459
2	1758	67.0	1758	89	AF134303
C 3	812	30.9	151580	82	AL590009
C 4	812	30.9	192017	82	AL590549
5	636	24.2	143812	66	AC021847
C 6	167	6.4	143812	66	AC021847
7	167	6.4	192017	82	AL590549
8	116	4.4	62537	78	AC090598

|||||
Db 541 CTTTCGAAGTCTCAAAATCAAGACGAGCTTTTCGATGCAAGACTTTGCCTATTTC 600
Qy 601 cattagagagacgtacgtgtttgtaacagcctccactctcaataactcaactcctt 660
|||||
Db 601 CATTACAGAGAGTACGATGTTGTGAACAGCCTCCACTCTCAATATACACCTCTT 660
Qy 661 atagagatgatgtaagaaggtctgaagttttatataccaatcctctgtatttttgatc 720
Db 661 ATAGAGATGATGTAAGAAGGTCTGAAGTTTATACCAATCTTCGTATTCTTTGATC 720
Qy 721 tatggaagaaaaatgttgcagaatcacagagataagaaggaagaaagagagacaga 780
Db 721 TATGGAAGAAAAATGTTCAGATACAGAGGATGAAGAGGAAGAAAGAGGAAGCAGA 780
Qy 781 agcagaaaaatctagatcgtctctcatgaaccagaaaaagtgcgaagagcacctcatgaca 840
Db 781 AGCAGAAAAATCTAGATCGTCTCTCATGACCAAGAAAAAGTGCCACAGCACCTCATGACA 840
Qy 841 ggccggagaatggcagaagctggcccaaggtccagagctggctgaagatgatgctaate 900
Db 841 GCGCGCAGAAATGGCAGAGCTGGCCCAAGTCCAGAGCTGGCTGAAGATGATGCTAATC 900
Qy 901 tcttacataagaatattgaagttgcttaagtggccagcctctcaattttgaaacaagacctc 960
Db 901 TCTTACATAAGCATATGTGAAGTTGCTTAATGGCCAGCCTCTCATTTTGAACAAGACCTC 960
Qy 961 agacatacgtggatcatacgtatggatcttactcactcttctgcttgcctattagtcaga 1020
Db 961 AGACATACGTGGATCATATGGATGGATCTTACTCACTTTCTGCTTGGCCATTTAGTCAGA 1020
Qy 1021 tgaagtgcctctgactagagctgaggaagggtatttagtcagacacacatgaaccacctc 1080
Db 1021 TGAGTGAGCTTCTGACTAGAGCTGAGGAAGGCTATTAGTCACACACATGAACCACTC 1080
Qy 1081 caactccaccaatgcattgagcaggagatgcataaacatgatacccaactgtatcagttctg 1140
Db 1081 CACTCTCCACCAATGCAATGGAGCAGAGAGATGCAAAACCGATACCCACTGTATCAGTTCTG 1140
Qy 1141 ctacagtttgatagaaatcgccctcagtcaccagctcagcagcagcagcagcagctgtttg 1200
Db 1141 CTACAGTTTGATAGAAAATCGCCCTCAGTCACAGCTACAGCAGCAACACCTGTGTTTG 1200
Qy 1201 tgaagcccaactcccaactctcccaactctctccactctctcactctgtcactctcact 1260
Db 1201 TGAGCCCACTCCCACTCCCTCCCACTCTTCCATCTGCTTGTCAACTTCTCAT 1260
Qy 1261 taagagcttcaatgaactcaactcctccctccctccagtaacctcccccaactccactccag 1320
Db 1261 TAAGAGCTTCAATGACTTCAACTCTCCCTCCCACTTACCTCCCACTCCCACTCCAG 1320
Qy 1321 ccaactgtttggaagctccagcagtagtaccacacacacacacacacacacacacacacac 1380
Db 1321 CCACTGCTTGGAGCTCCAGAGTAGTACCACCACTCCAGCTCCTCTTCAGATTGCCCTTG 1380
Qy 1381 gagtcttcaaccagctcctctcccaattgacacctcctctctagtacagccctctccaccag 1440
Db 1381 GAGTTCTCACCCAGCTCCCTCCCTCCCAATTGCACCTCTCTAGTACAGCCCTCTCCACAG 1440
Qy 1441 tagctagagctgccccagtagtgagactgtacagttacagttcaactcccaacaaaggtgaag 1500
Db 1441 TAGCTAGAGCTGCCAGTAGTGAGACTGTACCAAGTTTACCACTTCCCACTCCCAAGAGTGAAG 1500
Qy 1501 ttcaagggtgcttccacccccac 1560
Db 1501 TTCAAGGGGCTGGCTCCACCCCACTCCCTCTCTCTGCTCCCTCCCTCCCTCCCTCCCTCC 1560
Qy 1561 catacctgtcacagttacagctcttctcctcctcctcctcctcctcctcctcctcctcctc 1620
Db 1561 CATCACCTGTACAGTTTACAGCTCTTGTCTCATCTCCTCTGCTGGCTACATCCCACTCCAT 1620
Qy 1621 ctactgcccaggtccccatgttccattaatgctccatcctcctcctcctcctcctcctcctc 1680
|||||

Db 1621 CTACTGCCCCAGGTCCCATGTTTCCATTAAATGCCTCCATCTCTCTCATCACAGTTATAC 1680
Qy 1681 ctgcttctgagccaaagcgcctcaatcaacccctacactgtaactagtgatgcccagagtg 1740
|||||
Db 1681 CTGCTTCTCAGCCAAAGCGCCATCCATCAACCTACCTGCTAAATCAGTGTATGCCAGAGTG 1740
Qy 1741 tgcctactgaaacaaatacgaagaggtattcagctacgaagtagaagagcagcgtgaac 1800
Db 1741 TGCTACTGGAAGCAATACGAAAAGGTATTTCAGCTAGCAAAAGTAGAAGACGACGCGAAC 1800
Qy 1801 aggaagctgaagcatgaacgcattgaaaaagatgttgcaccactcctgtctcgcgcgtatcg 1860
Db 1801 AGGAAGCTAAGCATGAACGCATTTGAAAACGATGTGCCACCACTCCTCTCCTGCCGTATTG 1860
Qy 1861 ctgttgaatatagtgattcgggaagatgattcagaatttgcagaatttgcagaatgattggttggagt 1920
Db 1861 CTGTTGAATATAGTGATTTCGGAAGATGATTACAGAAATTTGATGAAGTAGATTGTTGGAGT 1920
Qy 1921 agaaaaatgcattgtgataaataattacaaaactgaatgcataatgcataatgcataatgc 1980
Db 1921 AAGAAAAATGCTATGATAAATATTACAAAACCTGAATGCAAAATGCTCTTGTGTGCTTGT 1980
Qy 1981 tcttgaaaaatgtttgtctcattctagtttgccttcttcttcttcttcttcttcttcttctt 2040
Db 1981 TCTTGAAAATGTTTGGTCATCTAGTGTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2040
Qy 2041 tttctcctcaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2100
Db 2041 TTTTCTCTCATTAATTTTCTAAGGAAATATTAGCATACATTTCAAACTAAATGT 2100
Qy 2101 ttacagtggtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2160
Db 2101 TTTACAGTGGCTTATCTTTTCTCCCTCGAAAGACTAAATTTGGTCAAAATAAACCACT 2160
Qy 2161 aagttataagcatggacagctgtttagagtagcagattcagtttttcttcttcttcttctt 2220
Db 2161 AAGTATTAAGCATGGACAGCTGTTGTAGTAGTAGAGATTTCAGTTTTTGTATATATCTTA 2220
Qy 2221 attgttactcttctgaaatttcttcttcttcttcttcttcttcttcttcttcttcttctt 2280
Db 2221 ATTGTGCTACTTGTGAATTTTAATTTAAGAAAGCAACTGAAATTTGAAATCTTTGAGGCA 2280
Qy 2281 gctgtatctactaatgagccttattcctcctgtagtcttcttcttcttcttcttcttcttct 2340
Db 2281 GCTGTATCTACTAATGAGCCTTATTCATTTCTCTGATGTTTTTAAAAAGAGAAACACTGCC 2340
Qy 2341 ttgattatacgaatacactcagaagatcaatttagcttctgtagtcttcttcttcttcttct 2400
Db 2341 TTGATTATACGAATACACTCAGAAAGTACATTTAGCTTTGTAGTGTGGAATCTCTTAAG 2400
Qy 2401 gaatgcttgaatttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2460
Db 2401 GAATGCTGAATTTTTCATTTATTTTATTTTATTTATATCTTTCCTTATTTGAATG 2460
Qy 2461 tttagcagtagtccccctccccacttatattctgtagtcttcttcttcttcttcttcttct 2520
Db 2461 TTTAGCAGTATCCCTTCCCACTTATATTTGTGTATGATTGTTTCTTCTTCTTCTTCTT 2520
Qy 2521 gttaaaaacttttccatgtgaaatactctgacttaaacatacacatgtaactacataactg 2580
Db 2521 GTTAAAAAATTTTCCATGTGAAATACTCTGACTTTAAACATACATGTAACATTACATACTG 2580
Qy 2581 ttaagaataacagctctgatttaataaagtgttctatttttaaaagt 2625
Db 2581 TTAAGAATAACAGCTCTGATTTAATAAATGTTTCATTTTAAAAAGTT 2625
|||||

RESULT 2

AF134303

LOCUS

DEFINITION

ACCESSION

VERSION

AF134303

Homo sapiens

Scarl (K1AA0269)

mRNA, complete cds.

1758 bp

PRI

14-JUN-1999

GI:4927209


```
|||||
Db 100779 TCTAAGGAAAATATTAGCATACATTCAAACTAAATGTTTACAGTGGCTTATCTTTT 100720
QY 2123 tccccccgaaagactaatttggtcaataaacacactaagtattaaacagacagctg 2182
|||||
Db 100719 TTCCTCCCTGAAAGAACTAATTTGGTCAATAAACCACTAAGTATTAAAGCATGGACAGCTG 100560
QY 2183 ttgttagagtagcagattcagttttttgatatatcttaattgtgtactttgtgaatttta 2242
|||||
Db 100659 TTGTTAGAGTACGACATTCAGTTTTTTGATATATCTTAATCTGTACTTTGTGGAATTTTA 100600
QY 2243 atttaaaagaacactgaataattgaaattgtgagggcagctgtatctactaatgagcctt 2302
|||||
Db 100599 ATTTAAGAAAGCAACTGAAATTTGAAATCTTGAGGCGAGCTGTCTACTAATGAGCCTT 100540
QY 2303 attcattctcctgatgttttaaaagaacacactgccttgattatagaaatacactcag 2362
|||||
Db 100539 ATTCCATTTCCCTGATGTTTAAAGAAGAAACACATGCGCTTGATTATAGCAATACACTCAG 100480
QY 2363 aaagtacatttagctgttagttgaattctcttaagaagaatgcttgaattttttcatta 2422
|||||
Db 100479 AAAGTACATTTAGCTGTAGTGTGAATCTCTTAAGGAATGCTTGAAATTTTTCATTA 100420
QY 2423 ttgttttattgttttatatacttccttattgttaattgttagcagtagtccccctccac 2482
|||||
Db 100419 TTGTTTTATTGTTTTATATACTTGCCTTATTTGAATGTTTAGCAGTATCCCTTCCAC 100360
QY 2483 ttatatattgtgatgatatttcttctgcctataggagttaaaacitttccatgtgaa 2542
|||||
Db 100359 TTATATATTGTGTGATATGATTTTCTGCTGCTTATAGGAGTTAAACATTTTCCATGTGAA 100300
QY 2543 atactctgacttaaacatcacatgtaacttacataactgttaagaataacagctgtgattta 2602
|||||
Db 100299 ATACTCTGACTTAACATACATGTAACCTTACATATACTGTTAAGATATACAGTCTCATTTA 100240
QY 2603 ataaatggttcatttaaaagtt 2625
|||||
Db 100239 ATAAATGGTTCACTTTTAAAGTT 100217

RESULT 4
AL590549/c
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-403I19, 14-APR-2001
PROGRESS ***, 11 unordered pieces.
ACCESSION AL590549
VERSION AL590549.3 GI:13660901
KEYWORDS HTG; HTGS_PHASE1.
SOURCE
  human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 192017)
  Slms, S.
  Direct Submission
  Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerquests@sanger.ac.uk
  On Apr 17, 2001 this sequence version replaced gi:13625115.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  ----- Project Information
  Center project name: BA403I19
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Consensus quality: 187117 bases at least Q40
  Consensus quality: 188982 bases at least Q30
  Consensus quality: 190116 bases at least Q20
```

```
Insert size: 191017; sum-of-contigs
Insert size: 195674; agarose-fp
Quality coverage: 4.85x in Q20 bases; sum-of-contigs Quality
coverage: 4.74x in Q20 bases; agarose-fp
-----
```

```
* NOTE: this is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 4711: contig of 4711 bp in length
* 4712 4811: gap of 100 bp
* 4812 15590: contig of 10779 bp in length
* 15591 15690: gap of 100 bp
* 15691 80383: contig of 64693 bp in length
* 80384 80483: gap of 100 bp
* 80484 126461: contig of 45978 bp in length
* 126462 126561: gap of 100 bp
* 126562 129107: contig of 2546 bp in length
* 129108 129207: gap of 100 bp
* 129208 148782: contig of 19575 bp in length
* 148783 148882: gap of 100 bp
* 148883 162967: contig of 14085 bp in length
* 162968 163067: gap of 100 bp
* 163068 170242: contig of 7175 bp in length
* 170243 170342: gap of 100 bp
* 170343 177820: contig of 7478 bp in length
* 177821 177920: gap of 100 bp
* 177921 184396: contig of 6476 bp in length
* 184397 184496: gap of 100 bp
* 184497 192017: contig of 7521 bp in length.
```

FEATURES

```
source
  1. 192017
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /clone_lib="RPC1-11.2"
    /clone="RP11-403I19"
  1. 4711
    /note="assembly_fragment:01747"
    /fragment_chain:1"
  4812..15590
    /note="assembly_fragment:00630"
    /fragment_chain:1"
  15691..80383
    /note="assembly_fragment:01340"
    /fragment_chain:1"
  80484..126461
    /note="assembly_fragment:01664"
    /fragment_chain:1"
  126562..129107
    /note="assembly_fragment:01855"
    /fragment_chain:2"
  129208..148782
    /note="assembly_fragment:02190"
    /fragment_chain:2"
  148883..162967
    /note="assembly_fragment:02427"
    /fragment_chain:2"
  163068..170242
    /note="assembly_fragment:01301"
    /fragment_chain:2"
  170343..177820
    /note="assembly_fragment:00226"
    /fragment_chain:3"
  177921..184396
    /note="assembly_fragment:00479"
    /fragment_chain:3"
  184497..192017
    /note="assembly_fragment:01866"
```

fragment_chain:3"																						
BASE COUNT	57883 a	37927 c	37166 g	58039 t	1002 others																	
ORIGIN																						
Query Match 30.9%; Score 812; DB 82; Length 192017;																						
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;																						
Matches 862; Conservative 0;																						
QY	1763	aggtattcagctacgcaaaagtagaagacagcgtgacacgaggaagctaaagcaacgcatt	1822																			
Db	118245	AGGTATTACGTACGCAAAAGTAGAAGACAGCAGCTGAACAGGAGCTAAGCATGAACGCAT	118186																			
QY	1823	tgaacacgattgtgccaccatcctgctccgctattgtgttgatataagtgattcggaa	1882																			
Db	118185	TGAAAACGATGTTGCCACCATCTGCTCCGCTATTGCTGTGTAATAGTATTCGGA	118126																			
QY	1883	agatgattcagaattgatgaagtagattggttgagtagaagaaaaatgattgataaata	1942																			
Db	118125	AGATGATTTCAGAAATTTGATGAAGTAGATTGGTTGGAGTAGAAAAATGATTGATAAATA	118066																			
QY	1943	ttacaaaactgaatgcaaatgtcctttgttggtgctgttccttgaaaaatggttggtcatt	2002																			
Db	118065	TTACAAAACCTGAATGCAAAATGTCTTTGTGGTGTGTTCTTGTGAAAATGTTTGGTCATT	118006																			
QY	2003	ctagtgttttgccttttcccttataataaagaccccttttccctccataacttttgatt	2062																			
Db	118005	CTAGTGTTTTGTCTTCTTTTCCCTTATAATAAATGACCCCTTTCTCTCCATAAATTTGATT	117946																			
QY	2063	tctaagaaaaattagcatcatttcaactcaaatggttttccagtggttattctttttt	2122																			
Db	117945	TCTAAGAAAAATATTAGCATACATTTCAAACTAAATGTTTACAGTGGCTATTCCTTTTTT	117886																			
QY	2123	ttcccctgaaagactaaatttggtgcaataaaaccactaagtattaaagcatggacagctg	2182																			
Db	117885	TTCCCCCTGAAAAGACTAATTTGCTCAATAAACCACTAAGTATTAAAGCATGGACAGCTG	117826																			
QY	2183	tgttgatgtagcagattcagttttttgatatacttaattgtgacttttggaatttta	2242																			
Db	117825	TTGTTAGAGTAGCAGATTTCAGTTTTTTTGATATATCTTAATTTGTGTACTTTTGAATTTTA	117766																			
QY	2243	atttaagaaagcaactgaaatgaaatcttgaggcagctgtatctactaatgagccct	2302																			
Db	117765	ATTTAAGAAAGCACTGAATTTGAAATCTTGAGGCAGCTGTCTACTTAATGAGCCCT	117706																			
QY	2303	attccatttctcgtatgttttaaagaagaacacactgccttgattatacgaatacactcag	2362																			
Db	117705	ATPCCATTTCTCGATGTTTAAAGAAGAAACACACTGCCITGATTATACGAATACACTCAG	117646																			
QY	2363	aaagcacatttagcttgtagtgtagtgaattctctctaaaggaatgcttgaaatttttcatta	2422																			
Db	117645	AAAGTACATTTAGCTTGTAGTGTGAATTTCTCTTAAAGGAATGCTTGAATTTTTCATTA	117586																			
QY	2423	tgttttattgtttttatatactgcttattgaaatggttgaagtagatcccttcccac	2482																			
Db	117585	TTGTTTATGTTTTTATATACGTGCTTATTGTAATGTTTACAGTATCCCTTCCCCAC	117526																			
QY	2483	ttatatattgtgtgatgatgtttgcttgccatagaggttaaacttttccatgtgaa	2542																			
Db	117525	TTATATATTGTGTGATGATGATTTTGTCTGCTATAGAGTTTAAACTTTTCCATGTGAA	117466																			
QY	2543	atactctgacttaaacatacatgtaacttacataactgtttaagaataaacagctgtatta	2602																			
Db	117465	ATACTCTGACTTAAACATACATGTAACCTACATAAATGTTTAAAGAAACAGCTCTGATTTA	117406																			
QY	2603	ataaatgggttcattttaaaagtt	2625																			
Db	117405	ATAAATGGTTCATTTTAAAAGTT	117383																			
RESULT	5																					
AC021847																						
LOCUS	AC021847	143812 bp	DNA	HTG	07-JUL-2000																	

Homo sapiens chromosome 6 clone RP11-181P4, WORKING DRAFT SEQUENCES									
32 unordered pieces.									
ACCESSION	AC021847								
VERSION	AC021847.6	GI:8954243							
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
AUTHORS	1 (bases 1 to 143812)								
TITLE	Waterston,R.H.								
JOURNAL	The sequence of Homo sapiens clone								
AUTHORS	Unpublished								
TITLE	2 (bases 1 to 143812)								
JOURNAL	Waterston,R.H.								
AUTHORS	Direct Submission								
TITLE	Submitted (20-JAN-2000) Genome Sequencing Center, Washington								
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA								
COMMENT	On Jul 7, 2000 this sequence version replaced gi:8567974.								
----- Genome Center -----									
Center: Washington University Genome Sequencing Center									
Center code: WUGSC									
Web site: http://genome.wustl.edu/gsc/index.shtml									
----- Project Information -----									
Center project name: H_NH0181P04									
----- Summary Statistics -----									
Sequencing vector: M13; 100%									
Sequencing vector: plasmid; 0%									
Chemistry: Dye-primer ET; 100% of reads									
Chemistry: Dye-terminator Big Dye; 0% of reads									
Assembly program: Phrap; version 0.990319									
Consensus quality: 125690 bases at least Q40									
Consensus quality: 130641 bases at least Q30									
Consensus quality: 133684 bases at least Q20									
Insert size: 151000; agarose-fp									
Insert size: 140712; sum-of-contigs									
Quality coverage: 2.84 in Q20 bases; agarose-fp									
Quality coverage: 3.15 in Q20 bases; sum-of-contigs									

* NOTE: This is a 'working draft' sequence. It currently									
* consists of 32 contigs. The true order of the pieces									
* is not known and their order in this sequence record is									
* arbitrary. Gaps between the contigs are represented as									
* runs of N, but the exact sizes of the gaps are unknown.									
* This record will be updated with the finished sequence									
* as soon as it is available and the accession number will									
* be preserved.									

*	1	1345:	contig of 1345 bp in length						
*	1346	1445:	gap of unknown length						
*	1446	2816:	contig of 1371 bp in length						
*	2817	2916:	gap of unknown length						
*	2917	4030:	contig of 1114 bp in length						
*	4031	4130:	gap of unknown length						
*	4131	6512:	contig of 2382 bp in length						
*	6513	6612:	gap of unknown length						
*	6613	8567:	contig of 1955 bp in length						
*	8568	8667:	gap of unknown length						
*	8668	10333:	contig of 1666 bp in length						
*	10334	10334:	gap of unknown length						
*	10434	12876:	contig of 2443 bp in length						
*	12877	12976:	gap of unknown length						
*	12977	15219:	contig of 2243 bp in length						
*	15220	15319:	gap of unknown length						
*	15320	17854:	contig of 2535 bp in length						
*	17855	17854:	gap of unknown length						
*	17955	20526:	contig of 2572 bp in length						
*	20527	20626:	gap of unknown length						
*	20627	24354:	contig of 3728 bp in length						
*	24355	24454:	gap of unknown length						
*	24455	28206:	contig of 3752 bp in length						
*	28207	28306:	gap of unknown length						

Homo sapiens chromosome 6 clone RP11-181P4, WORKING DRAFT SEQUENCE, 32 unordered pieces.
AC021847
AC021847.6 GI:8954243
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143812)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 143812)
Waterston,R.H.
Direct Submission
Submitted (20-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8567974.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0181P04
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 125690 bases at least Q40
Consensus quality: 130641 bases at least Q30
Consensus quality: 133684 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 140712; sum-of-contigs
Quality coverage: 2.84 in Q20 bases; agarose-fp
Quality coverage: 3.15 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 1345: contig of 1345 bp in length
* 1346 1445: gap of unknown length
* 1446 2816: contig of 1371 bp in length
* 2817 2916: gap of unknown length
* 2917 4030: contig of 1114 bp in length
* 4031 4131: gap of unknown length
* 4131 6512: contig of 2382 bp in length
* 6513 8667: gap of unknown length
* 8667 10333: contig of 1955 bp in length
* 10333 10434: gap of unknown length
* 10434 12876: contig of 2443 bp in length
* 12876 15219: contig of 2243 bp in length
* 15219 15320: gap of unknown length
* 15320 17854: contig of 2535 bp in length
* 17854 17955: gap of unknown length
* 17955 20526: contig of 2572 bp in length
* 20527 20627: gap of unknown length
* 20627 24354: contig of 3728 bp in length
* 24355 24454: gap of unknown length
* 24455 28206: contig of 3752 bp in length
* 28207 28306: gap of unknown length

QY 1552 ttgacacataacacgtgtacacagttacagctcttgcctacatccctccctctgggctacatc 1611
|||||
Db 141587 TTCGACCATCATCACCTGTACAGTGTACAGCTTGTGCTCATCTCCCTCTGGGCTACATC 141646
|||||
QY 1612 caactcactactgccccaggtcccccatttccatttaattgctccatctctccatcac 1671
|||||
Db 141647 CAACTCCATCTACTGCCAGGTCCTCCATGTTCCATTAATGCTCCATCTCCCTCCATCAC 141706
|||||
QY 1672 aagtatacctgtcttgagcgaagcgcacatccatccacccctacctgtaataatcagtgatg 1731
|||||
Db 141707 AAGTTATACCTGCTTGTGAGCCAAAGCGCATCATCAACCTTACCTGTAATCAGTGATG 141766
|||||
QY 1732 ccaggagtgctgactggaagcaatacgaagagta 1767
|||||
Db 141767 CCAGGAGTGTGCTACTGGAAGCAATACGAAAGGTA 141802
|||||

RESULT 6
AC021847/c
LOCUS
DEFINITION AC021847 143812 bp DNA HTG 07-JUL-2000
Homo sapiens chromosome 6 clone RP11-181P4, WORKING DRAFT SEQUENCE,
32 unordered pieces.
ACCESSION AC021847
VERSION AC021847.6 GI:8954243
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143812)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 143812)
Waterston, R.H.
Direct Submission
Submitted (20-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8567974.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----
Center project name: H_NH0181P04

----- Summary Statistics -----
Sequencing vector: M13; 100%

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 125690 bases at least Q40

Consensus quality: 130641 bases at least Q30

Consensus quality: 133684 bases at least Q20

Insert size: 151000; agarose-fp

Insert size: 140712; sum-of-contigs

Quality coverage: 2.84 in Q20 bases; agarose-fp

Quality coverage: 3.15 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1345: contig of 1345 bp in length

* 1346 1445: gap of unknown length

* 1446 2816: contig of 1371 bp in length

* 2817 2816: gap of unknown length
* 2917 4030: contig of 1114 bp in length
* 4031 4130: gap of unknown length
* 4131 6512: contig of 2382 bp in length
* 6513 8567: contig of 1955 bp in length
* 8568 8667: gap of unknown length
* 8668 10333: contig of 1666 bp in length
* 10334 12876: contig of 2443 bp in length
* 12877 12976: gap of unknown length
* 12977 15219: contig of 2243 bp in length
* 15220 15319: gap of unknown length
* 15320 17854: contig of 2535 bp in length
* 17855 20526: contig of 2572 bp in length
* 20527 20626: gap of unknown length
* 20627 24354: contig of 3728 bp in length
* 24355 24454: gap of unknown length
* 24455 28206: contig of 3752 bp in length
* 28207 28306: gap of unknown length
* 28307 32673: contig of 4367 bp in length
* 32674 32773: gap of unknown length
* 32774 36150: contig of 3377 bp in length
* 36151 36250: gap of unknown length
* 36251 39810: contig of 3560 bp in length
* 39811 39910: gap of unknown length
* 39911 43789: contig of 3879 bp in length
* 43790 43889: gap of unknown length
* 43890 47197: contig of 3308 bp in length
* 47198 47297: gap of unknown length
* 47298 51028: contig of 3731 bp in length
* 51029 51128: gap of unknown length
* 51129 55156: contig of 4028 bp in length
* 55157 55256: gap of unknown length
* 55257 60671: contig of 5415 bp in length
* 60672 60771: gap of unknown length
* 60772 65410: contig of 4639 bp in length
* 65411 71071: contig of 5561 bp in length
* 71072 71171: gap of unknown length
* 71172 76415: contig of 5244 bp in length
* 76416 76515: gap of unknown length
* 76516 81312: contig of 4797 bp in length
* 81313 86835: contig of 5423 bp in length
* 86836 86935: gap of unknown length
* 86936 91557: contig of 4621 bp in length
* 91557 91657: gap of unknown length
* 91657 96940: contig of 5283 bp in length
* 96940 103443: contig of 6404 bp in length
* 103444 103543: gap of unknown length
* 103544 109673: contig of 6130 bp in length
* 109674 109773: gap of unknown length
* 109774 118660: contig of 8887 bp in length
* 118661 118760: gap of unknown length
* 118761 130272: contig of 11512 bp in length
* 130273 130373: gap of unknown length
* 130373 143812: contig of 13440 bp in length.

FEATURES

1..143812
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-181P4"
1..1345
/note="assembly_name:Contig10"
1446..2816
/note="assembly_name:Contig12"
2917..4030
/note="assembly_name:Contig14"
4131..6512

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature /note="assembly_name:Contig15"
6613. .8567
misc_feature /note="assembly_name:Contig17"
8668. .10333
misc_feature /note="assembly_name:Contig18"
10434. .12876
misc_feature /note="assembly_name:Contig19"
12977. .15219
misc_feature /note="assembly_name:Contig20"
15320. .17854
misc_feature /note="assembly_name:Contig21"
17955. .20526
misc_feature /note="assembly_name:Contig22"
20627. .24354
misc_feature /note="assembly_name:Contig23"
24455. .28206
misc_feature /note="assembly_name:Contig24"
clone_end:SP6
vector_side:right
28307. .32673
misc_feature /note="assembly_name:Contig25"
32774. .36150
misc_feature /note="assembly_name:Contig26"
36251. .39810
misc_feature /note="assembly_name:Contig27"
39911. .43789
misc_feature /note="assembly_name:Contig28"
43890. .47197
misc_feature /note="assembly_name:Contig29"
47298. .51028
misc_feature /note="assembly_name:Contig30"
51129. .55156
misc_feature /note="assembly_name:Contig31"
55257. .60671
misc_feature /note="assembly_name:Contig32"
60772. .65410
misc_feature /note="assembly_name:Contig33"
65511. .71071
misc_feature /note="assembly_name:Contig34"
71172. .76415
misc_feature /note="assembly_name:Contig35"
76516. .81312
misc_feature /note="assembly_name:Contig36"
81413. .86835
misc_feature /note="assembly_name:Contig37"
86936. .91556
misc_feature /note="assembly_name:Contig38"
91657. .96939
misc_feature /note="assembly_name:Contig39"
97040. .103443
misc_feature /note="assembly_name:Contig40"
103544. .109673
misc_feature /note="assembly_name:Contig41"
109774. .118660
misc_feature /note="assembly_name:Contig42"
118761. .130272
misc_feature /note="assembly_name:Contig43"
130373. .143812
misc_feature /note="assembly_name:Contig44"
41394 a 27942 c 28670 g 42684 t 3122 others
BASE COUNT
ORIGIN

Query Match 6.4%; Score 167; DB 66; Length 143812;
Best Local Similarity 100.0%; Pred. No. 1e-80;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 213 agdtgaactggcacaaagttaattcaagatgcqctagtgaaagaacatcgatcct 272
|||||
Db 50730 AGGTGAACGGCACAAAGGTTAATCTCAAGATGCGCTAGTGAAGAAGAACATCGATCCT 50671
Qy 273 aggcactgtgcccacacagcactgctgagggcattagaatgaactgaatgtgtacc 332
|||||

Db 50670 AGGCACCTTGCCACACAGCAGCAGCTAGAGGCAATTAAGAATGAAGTGTGTAAACC 50611
Qy 333 aatatttcttgccgaataataattagacaaactaacctaaagtaa 379
|||||
Db 50610 AATATTTCTTGCCAAATATATATTAGACAACTAAGTACGCTAAGTAA 50564
RESULT 7
AL590549 192017 bp DNA HTG 14-APR-2001
LOCUS Homo sapiens chromosome 6 clone RP11-403119, *** SEQUENCING IN
DEFINITION PROGRESS ***, 11 unordered pieces.
ACCESSION AL590549
VERSION AL590549.3 GI:13660901
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims.S.
Direct Submission
Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13625115.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA403119
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 187117 bases at least Q40
Consensus quality: 188982 bases at least Q30
Consensus quality: 190116 bases at least Q20
Insert size: 191017; sum-of-contigs
Insert size: 193674; agarose-fp
Quality coverage: 4.85x in Q20 bases; sum-of-contigs Quality
coverage: 4.74x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4711: contig of 4711 bp in length
* 4712 4811: gap of 100 bp
* 4812 15590: contig of 10779 bp in length
* 15591 15690: gap of 100 bp
* 15691 80383: contig of 64693 bp in length
* 80384 80483: gap of 100 bp
* 80484 128461: contig of 45978 bp in length
* 128462 126561: gap of 100 bp
* 126562 129107: contig of 2546 bp in length
* 129108 129207: gap of 100 bp
* 129208 148782: contig of 19575 bp in length
* 148783 148882: gap of 100 bp
* 148883 162967: contig of 14085 bp in length
* 162968 163067: gap of 100 bp
* 163068 170242: contig of 7175 bp in length
* 170243 170342: gap of 100 bp
* 170343 177820: contig of 7478 bp in length
* 177821 177920: gap of 100 bp
* 177921 184396: contig of 6476 bp in length
* 184397 184496: gap of 100 bp

FEATURES * 184497 192017: contig of 7521 bp in length.
source Location/Qualifiers
1. .192017
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="RPCI-11.2"
1. .4711
/note="assembly_fragment:01747
fragment_chain:1"
4812. .15590
/note="assembly_fragment:00630
fragment_chain:1"
15691. .80383
/note="assembly_fragment:01340
fragment_chain:1"
80484. .126461
/note="assembly_fragment:01664
fragment_chain:1"
126562. .129107
/note="assembly_fragment:01855
fragment_chain:2"
129208. .148782
/note="assembly_fragment:02190
fragment_chain:2"
148883. .162967
/note="assembly_fragment:02427
fragment_chain:2"
153068. .170242
/note="assembly_fragment:01301
fragment_chain:2"
170343. .177820
/note="assembly_fragment:00226
fragment_chain:3"
177921. .184396
/note="assembly_fragment:00479
fragment_chain:3"
184497. .192017
/note="assembly_fragment:01866
fragment_chain:3"
BASE COUNT 57883 a 37927 c 37166 g 58039 t 1002 others
ORIGIN

Query Match 6.4%; Score 167; DB 82; Length 192017;
Best Local Similarity 100.0%; Pred. No. 1e-80;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 aggtgaactggcacaaggttaactcaagatgccgctagtgaaaagaacatcgatcct 272
|||||
Db 173070 AGGTGAACGGCAAAAGGTTAATCTCAAGATCCGCTAGTGAAGAAAGAACATCGATCCT 173129

QY 273 aggcacttggcacaagcagcactgcctagaggcattagaatgaactggaatgtgaacc 332
|||||
Db 173130 AGGCACCTGTGTCACACAGACACATGCTAGAGGCAATTAAGAAATGAATGTAAC 173189

QY 333 aatatttccttgccaaataataatagacaactaagtagcctaagtaa 379
|||||
Db 173190 AATATTTCTTGGCAATATATAATTAGACAACTAAGTAGCCTAAGTAA 173236

RESULT 8
AC090598
LOCUS AC090598 62537 bp DNA HTG 25-MAR-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-55506 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC090598
VERSION AC090598.2 GI:13446291
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 62537)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Cammarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Coyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 25, 2001 this sequence version replaced gi:13194331.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11841
Center clone name: 555_O-6

* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 689: contig of 689 bp in length
* 690 789: gap of 100 bp
* 790 1505: contig of 716 bp in length
* 1506 1605: gap of 100 bp
* 1606 2322: contig of 717 bp in length
* 2323 2422: gap of 100 bp
* 2423 3132: contig of 710 bp in length
* 3133 3232: gap of 100 bp
* 3233 3943: contig of 711 bp in length
* 3944 4043: gap of 100 bp
* 4044 4774: contig of 731 bp in length
* 4775 4874: gap of 100 bp
* 4875 5588: contig of 714 bp in length
* 5589 5688: gap of 100 bp
* 5689 6416: contig of 728 bp in length
* 6417 6516: gap of 100 bp
* 6517 7290: contig of 774 bp in length
* 7291 7390: gap of 100 bp
* 7391 8146: contig of 756 bp in length
* *

```
* 8147 8246: gap of 100 bp
* 8247 8962: contig of 716 bp in length
* 8963 9062: gap of 100 bp
* 9063 9804: contig of 742 bp in length
* 9805 9904: gap of 100 bp
* 9905 10646: contig of 742 bp in length
* 10647 10746: gap of 100 bp
* 10747 11516: contig of 770 bp in length
* 11517 11616: gap of 100 bp
* 11617 12334: contig of 718 bp in length
* 12335 12434: gap of 100 bp
* 12435 13161: contig of 727 bp in length
* 13162 13261: gap of 100 bp
* 13262 13983: contig of 722 bp in length
* 13984 14083: gap of 100 bp
* 14084 14843: contig of 760 bp in length
* 14844 14943: gap of 100 bp
* 14944 15654: contig of 711 bp in length
* 15655 15754: gap of 100 bp
* 15755 16460: contig of 706 bp in length
* 16461 16560: gap of 100 bp
* 16561 17298: contig of 738 bp in length
* 17299 17398: gap of 100 bp
* 17399 18145: contig of 747 bp in length
* 18146 18245: gap of 100 bp
* 18246 18955: contig of 710 bp in length
* 18956 19055: gap of 100 bp
* 19056 19834: contig of 779 bp in length
* 19835 19934: gap of 100 bp
* 19935 20665: contig of 731 bp in length
* 20666 20765: gap of 100 bp
* 20766 21484: contig of 719 bp in length
* 21485 21584: gap of 100 bp
* 21585 22313: contig of 729 bp in length
* 22314 22413: gap of 100 bp
* 22414 23129: contig of 716 bp in length
* 23130 23229: gap of 100 bp
* 23230 23974: contig of 745 bp in length
* 23975 24074: gap of 100 bp
* 24075 24811: contig of 737 bp in length
* 24812 24911: gap of 100 bp
* 24912 25705: contig of 794 bp in length
* 25706 25805: gap of 100 bp
* 25806 26531: contig of 726 bp in length
* 26532 26631: gap of 100 bp
* 26632 27345: contig of 714 bp in length
* 27346 27445: gap of 100 bp
* 27446 28164: contig of 719 bp in length
* 28165 28264: gap of 100 bp
* 28265 28950: contig of 686 bp in length
* 28951 29050: gap of 100 bp
* 29051 29813: contig of 763 bp in length
* 29814 29913: gap of 100 bp
* 29914 30637: contig of 724 bp in length
* 30638 30737: gap of 100 bp
* 30738 31545: contig of 808 bp in length
* 31546 31645: gap of 100 bp
* 31646 32364: contig of 719 bp in length
* 32365 32464: gap of 100 bp
* 32465 33247: contig of 783 bp in length
* 33248 33347: gap of 100 bp
* 33348 34132: contig of 785 bp in length
* 34133 34232: gap of 100 bp
* 34233 35003: contig of 770 bp in length
* 35003 35102: gap of 100 bp
* 35103 35875: contig of 773 bp in length
* 35876 35975: gap of 100 bp
* 35976 36673: contig of 698 bp in length
* 36674 36773: gap of 100 bp
* 36774 37479: contig of 706 bp in length
* 37480 37579: gap of 100 bp
* 37580 38293: contig of 714 bp in length
* 38294 38393: gap of 100 bp
```

```
* 38394 39163: contig of 770 bp in length
* 39164 39263: gap of 100 bp
* 39264 39983: contig of 720 bp in length
* 39984 40083: gap of 100 bp
* 40084 40804: contig of 721 bp in length
* 40805 40904: gap of 100 bp
* 40905 41689: contig of 785 bp in length
* 41690 41789: gap of 100 bp
* 41790 42519: contig of 730 bp in length
* 42520 42619: gap of 100 bp
* 42620 43345: contig of 726 bp in length
* 43346 43445: gap of 100 bp
* 43446 44161: contig of 716 bp in length
* 44162 44261: gap of 100 bp
* 44262 44959: contig of 698 bp in length
* 44960 45059: gap of 100 bp
* 45060 45766: contig of 707 bp in length
* 45767 45866: gap of 100 bp
* 45867 46635: contig of 769 bp in length
* 46636 46735: gap of 100 bp
* 46736 47449: contig of 714 bp in length
* 47450 47549: gap of 100 bp
* 47550 48287: contig of 738 bp in length
* 48288 48387: gap of 100 bp
* 48388 49188: contig of 801 bp in length
* 49189 49288: gap of 100 bp
* 49289 50005: contig of 717 bp in length
* 50006 50105: gap of 100 bp
* 50106 50835: contig of 730 bp in length
* 50836 50935: gap of 100 bp
* 50936 51706: contig of 771 bp in length
* 51707 51806: gap of 100 bp
* 51807 52521: contig of 715 bp in length
* 52522 52621: gap of 100 bp
* 52622 53360: contig of 739 bp in length
* 53361 53460: gap of 100 bp
* 53461 54173: contig of 713 bp in length
* 54174 54273: gap of 100 bp
* 54274 55018: contig of 745 bp in length
* 55019 55118: gap of 100 bp
* 55119 55852: contig of 734 bp in length
* 55853 55952: gap of 100 bp
* 55953 56712: contig of 760 bp in length
* 56713 56812: gap of 100 bp
* 56813 57531: contig of 719 bp in length
```

Query Match 4.48; Score 116; DB 78; Length 62537;
Best Local Similarity 100.0%; Pred. No. 2.3e-52;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 cttcttgcacttgcgcatgaactggaataacgcatgaagaagacacatccgact 60
|||||
Db 46211 CTTCTTGCACCTTCCGGATGATGAACCTGGATACGATGAAGAAAGAACACATCCGATCT 46270
|||||
Qy 61 caacattcactcttgcctataaccgatttaattatgattcccccagctagactag 116
|||||
Db 46271 CAACATTCACGCTCTGCCCTATACCGATTAAATTAATGATCCCGAGTACACTAG 46326
|||||
```

RESULT 9

LOCUS	GI9661	human STS A001021	DNA	STS	24-JUL-1996
DEFINITION	human STS A001021, sequence tagged site.				
ACCESSION	GI9661				
VERSION	GI9661.1 GI:1340232				
KEYWORDS	STS; STS sequence; primer; sequence tagged site.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 225)				
AUTHORS	Adams, M.D.				
JOURNAL	Unpublished (1996)				

COMMENT

Contact: Mark Adams
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Email: mdadams@tigr.org

Primer A: CTTACGACTAGACAATATGAAG
Primer B: ACTCTGACTTAACATACATGTA
STS size: 225
PCR Profile:

Denaturation: 96C 5min
Anneal: 54C 30sec
Extend: 72C 30sec
Denature: 95C 30sec
FinalExtend: 72C 5min
Cycles: 30

Protocol: GenomicDNA: 25 ng
Primer: 0.43 uM each
dNTPs: 230 uM each
AmpliTaq: 0.5 units
TagStart Ab: 0.5 units
Total Volume: 10 ul

Buffer: Tris-HCl pH8.8: 100 mM
KCl: 500 mM
MgCl2: 20 mM
Trislon X-100: 1%
Concentration: 10X

Prepared with primer pairs derived from THC41826; GenBank Accession
Numbers-- F10585, T87835, M79145, T06478, T32453, L51665.

FEATURES

Source
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
STS
primer_bind 1..23
primer_bind complement(203..225)
BASE COUNT 82 a 32 c 31 g 80 t
ORIGIN

Query Match 3.1%; Score 81; DB 54; Length 225;
Best Local Similarity 100.0%; Pred. No. 4.3e-33;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2545 actctgacttaaacatcacatgttaacttaacataactgttaagaataacagtcgtatttaac 2604
|||||
Db 225 ACTCTGACTTAACATACATGTAACATACATACTGTTAAGAATAACAGTCTGATTTAAT 166
QY 2605 aaatggttcattttaaagt 2625
|||||
Db 165 AAATGTTTCATTTTAAAGTT 145

RESULT 10

AF290877 2469 bp mRNA ROD 10-OCT-2000
LOCUS Mus musculus WAVE-1 mRNA, complete cds.
DEFINITION AF290877
ACCESSION AF290877
VERSION AF290877.1 GI:9931545
KEYWORDS house mouse.
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2469)
AUTHORS Westphal,R.S., Soderling,S.H., Alto,N.M., Langeberg,L.K. and Scott,J.D.
TITLE Scar/WAVE-1, a wiskott-aldrich syndrome protein, assembles an actin-associated multi-kinase scaffold

JOURNAL MEDLINE
20428428
PUBMED 10970852
REFERENCE 2 (bases 1 to 2469)
AUTHORS Soderling,S.H., Westphal,R.S. and Scott,J.D.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2000) Howard Hughes Medical Institute/Vollum Institute, Oregon Health Sciences University, 3181 Sam Jackson Park Rd., Portland, OR 98201, USA

FEATURES Source
1..2469
/organism="Mus musculus"
/db_xref="taxon:10090"

CDS

1..1680
/note="kinase-anchoring protein; Abl-binding protein;
Arp2/3 complex activator"
/codon_start=1
/product="WAVE-1"
/protein_id="AAG02214.1"
/db_xref="GI:9931546"
/translation="MPLVKRNIDPRHLCHTLPRGINKLECVTNLSANIIRQLSSL
SKYAEIDFGLFNEAHFSFRVNSLQERVDRLSVTQDPKEELSQDITMRKAFR
SSTIQDQLFDRKTLPILOETVDCEQPPNLTPTVRRDCKEGLKFTNPSYFDL
WKEMLQDTEDEKREKQKOKNDRPHEKVPVPRHRRERWOKLAGGPELAEEDA
DLHRHIEVANGPASHYETRPQTYVDHMDGSLSALPFSQMSSELLTRAERVLVRPH
EPPPPPMHGADGDAKPTCTISSATGLIENRPPQPAAGTTFVFSPTPPPPPLPSA
LSTSLRASMTSTPPPPVPVPPPPATLQAPVPPPPAPLQIAPGLHPAPPPPIAPP
LVOPSPVARAAVCETVPHPLPOGEVGLPPPPPPPPPPPPGIRPSPSPVAAALAH
PESGLHPASTAGPHALMPPSPSOVLPASEKRPSTLPVSDARSLLEATRKQ
IOLRKEQREGEAKHERIENDVAILLSRIAEVSDSDSEDFEVDWLE"

BASE COUNT 659 a 638 c 521 g 650 t 1 others
ORIGIN
Query Match 1.8%; Score 46; DB 94; Length 2469;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1180 caggcagaacacctgtgtttgagcccaactccccacactctcc 1225
|||||
Db 938 CAGGCAGAACACCTGTGTTGTGAGCCCACTCCCCACCTCCTCC 983

RESULT 11

AC019100 213475 bp DNA PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-443K8 from 2, complete sequence.
DEFINITION AC019100
ACCESSION AC019100
VERSION AC019100.4 GI:9454638
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 213475)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 213475)
AUTHORS Kalicki,J., Drone,K. and Belter,E.
TITLE The sequence of Homo sapiens BAC clone RP11-443K8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 213475)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 213475)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2000) Genome Sequencing Center, Washington


```

repeat_region 28931..29020
/rpt_family="MERL_type"
repeat_region 30276..30528
/rpt_family="MIR"
repeat_region 31554..31743
/rpt_family="MERL_type"
repeat_region 31791..31900
/rpt_family="MIR"
repeat_region 32232..32458
/rpt_family="MIR"
repeat_region 33120..33416
/rpt_family="Alu"
repeat_region 33715..33845
/rpt_family="ERV1"
repeat_region 34001..34230
/rpt_family="ERV1"
repeat_region 34482..34533
/rpt_family="L2"
repeat_region 34565..34752
/rpt_family="MIR"
repeat_region 34860..35000
/rpt_family="L2"
repeat_region 35455..35698
/rpt_family="MIR"
repeat_region 36250..36606
/rpt_family="MALR"

Query Match 1.0%; Score 25; DB 87; Length 213475;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1065 ccacatgaaccacctccacccacac 1089
|||||
Db 128388 CCACATGAACCCCTCCACCTCCAC 128412

RESULT 12
AB010070/c
LOCUS AB010070 78172 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MBK20.
ACCESSION AB010070 BA000015
VERSION AB010070.1 GI:2760166
KEYWORDS
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MBK20.
ORGANISM
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (sites)
AUTHORS Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyaajima,N.
and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. IV.
Sequence features of the regions of 1,456,315 bp covered by
nineteen physically assigned P1 and TAC clones
JOURNAL DNA Res. 5 (1), 41-54 (1998)
MEDLINE 98290546
REFERENCE 2 (bases 1 to 78172)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MBK20
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremmlin2ool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is T2I1 and the 3' clone is MXM12.

FEATURES
source
1..78172
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MBK20"
/clone_lib="Mitsui P1"
214..837
gene_id:"MBK20.1"
note="emb|CAB87947.1"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAB11436.1"
/db_xref="GI:9759579"
/translocation="MASFEESDLAIOASHLLELLVCDGFMGDFDASFYVGLMCI
EPHVPKQDPSPVLPDPSFVNEFLOVEGESSSSSPELNSSSTVETDQSVKARFE
EYDARHVGVRPRPWGFAAEIRDPAPKAGSRIMWLTGTFESDYDARAYDCAAFKLGR
KAVLNPPLDAGYAPANSRRKRSDDVHEEQLRTQSNSSSSCCDAF"
complement(join(2200..2567,2679..2786,2884..2956,
3043..3219,3367..3855,4103..4219))
note="gene_id:MBK20.2"
/codon_start=1
/evidence=not_experimental
/product="WD-repeat protein-like"
/protein_id="BAB11437.1"
/db_xref="GI:9759580"
/translocation="MEFTAYKQTGCPFCSPNSRYVAVANDYRLVIRDFSPQVOLF
SCLDKISYIEWALDSEYILCGYKKPMIOAWSLTOPWTKIDEGPAGISYARWSPDS
RHILTSEQLRLTVKSLNTACVHVPKHKGVSENOCKFAAICTRRCKDYVN
LLSCQSWEIMGSFVDTLDLADLEWSPDSSIVWDSPLYEYKAYEGCLGVKTVSASP
GQFLAIGSYDQMLRVNLHMTFAEFLHLSVRAPCSAAIKFEIDEPQLDMSLSL
DENFMSYDASEGYSIVRYEYMELPVAFPPQKPADKPNKQGVGLLAWSKDSQYIC
TRNDSMPTALWIMRLLEVAAILVQKEIPRAAVMDPTKRLILCTGSHLYMWTPSG
AFVCNPLPGFSIDLKWNIDGSCLLDKDKAFCCATVPSLPESDYSDD"
complement(join(5140..5587,5725..5969))
note="gene_id:MBK20.3"
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11438.1"
/db_xref="GI:9759581"
/translocation="MFSFLIFLELVKVVIAVVASIVFVFGSLILAGTAVGLTVTTP
LFIFSPILVPATIAITLTGTTGGALGATAIALIRMGVKSNNIPALCAPTMM
FAQSLTPIKINYEYFGKSGWKGKSPQATPNFYSGYGTWANNMGSTFKGFDQSGGG
STAGGSTPEAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
TGKAGKAGSKKK"
6556..7818
note="emb|CAB85517.1"
gene_id:"MBK20.4"
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11439.1"
/db_xref="GI:9759582"
/translocation="MSSCSRTTKAPRSRRNGGFSSSSATIVADIDVLIQISLF
LPKILLRKRYSKRWLSLITNPVSNRVKSNHPLPISGFFLHSPREIKYSEVSLDD
DATNORISSSLPWFTDQHTDMLIMQSTNGLLCKSCASSNHNTYVYVYPTTKQY
TLLHQIAGHIALSLAFDPSRSPHYKVCULGRSNSSSSASDSSELYHIIVYSSNGLW
RRVVPVTPSPFTFIETFSYFWNGAVNMGYFSSRDCLSFDTINTKILPLPDHHEH
EPLPDPRILMFLDESQGNLYYTEVNNQSSNLRVYEMESNSSWSYKYNVDLEPLAA

```

PEMIRVEYTRDRRIYAFVIGFVKKEETDAASYILLHLPNOAVKYNFIDTKFFKKLCDF
KSLVNDAPEDHFYRPORTFOFIKSLANV"
complement(join(7889..8469,9056..9554))
/note="contains similarity to receptor-like protein
kinase
gene_id: MBK20.5"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11440.1"
/db_xref="GI:9759583"
/translation="MRSKLLTLLGVLSAFLIVIFFIIFLRKESSTESDDYDVE
VLNRKQFSETSELVIFOGGDLTICIDAPGEVIGSKSYGLYKASLQRSGKIR
SLRFLPVPCTVRSQKFGFIETLGFVRHENLVLPLGFIAGNKGELMVHPFGSGN
LSDFIRSGDDSRKWINLRITIGISKALDHLHTGMOKPIVHGKLSKNVLLSSFFP
RISDFGHLNLSAGEILDVSAEGYKAPELIKMDVSKESDVYSLGVIMLELVS
KEPINEATGDEYFLDFMRNAVLDHRLSDLYRPEILSGDDNLSEECVLKYFOLAMS
CCSPSPSLRPNVQVLRKLEIGF"
11636..11722
/note="gene_id:MBK20.6"
/evidence=not_experimental
/product="cRNA-Leu(CAA)"
join(12520..12539,12603..12763,12892..13043,13128..13246,
13336..13447,13592..13681,13826..13897,14025..14162,
14266..14421,14707..14845,14945..15011)
/note="gb|AAD38624.1
gene_id: MBK20.7
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11441.1"
/db_xref="GI:9759584"
/translation="MQFALPDGPESEENVTRILKVAWTVPLGIATIAACIFVLW
ONLSYDPAQAIWIFGACVLEMAEPLXILSOTLLKQLRLIVETVATLCVTLV
SLLMQNMEXVIFALISQVAYGSLGFIYWAYFLICGVYRSSNLFPRFGNPFDFN
DLRMCKLFTQSKRLQVLTQGLVPLPGSLVVMVFLPFEESSYTFIFARFAGDQERN
KLGLIYVAUKVLILGLIFMAFGPSYSYLRLYGERWSDGEASLALQFYCLYII
VLAMNGTSEAFHVAVGKNELESDMLIFSLIYVALNILLIRSAAGIILMANSL
WQILLSIIIFLISEKLTILDRKNFETPLHFGIVGICFCLSAIVIVRRERVKRIK
RFRDYNHDD"
join(15335..15610,15690..16364)
/note="contains similarity to RING zinc finger protein
gene_id: MBK20.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11442.1"
/db_xref="GI:9759585"
/translation="MCVSTFWLYCKGLVSEVIRDESQIGGFGVACIDHEDNRLVE
MNVILGGEESTPHQAAELALIHANWALELDLGRVTFCCDSNILEYTVGKAEPNES
TVATLVREVSLLQSKFCEALPYMKDITFVILKAKDAIASQIRREGDYMETCPVC
YEHVTSDEKFEVPGCFHRCFDCIKKQADVALEAFKVPVNCPSFGCNSQLQRECEGV
LKPOLDMTWYKKASMKAKVLDVCCCTCDNVMAKPDLEYTKTFVDAELSGVRK
CTEGYCGFCGCRAGHSGMTCVEYFKRESNEPSPEDVR"
complement(join(16863..16984,17088..17154,17272..17340,
17462..17578,17642..17713,17790..17884,18056..18137,
18232..18282,18422..18574,18643..18684,18782..18880,
18966..19037,19148..19216,19525..19593,19716..19832,
19908..19979,20055..20122,20745..20826,20904..20954,
21053..21133,21251..21292,21434..21532,21630..21734,
21845..21917,22031..22302))
/note="gene_id:MBK20.9
pir|T17454
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11443.1"
/db_xref="GI:9759586"
/translation="MSLVEISGSMAAMPGRVPPPPPPRRLPMPDFADHFT
GAGWVGFRPAKKRSLKPLHWYKITSDQGSWLDQLRHQSDTAIEDISELET
LFVFEAKPERILHDLRASYRVFNRYMANKVNIINLSMPLMDPMLVANDSEV
VDVQIKLTKFCPTNDEMELLTYTGDKAALGYEOLLELMKVPRLKALRVFSFK
TOFCTKTTELKRLNVTSACEVRSSEKLKEMKKIPCLGNTSNQGPDRGVKLDSS
VSDTHYVSHHYCKVLASASELLDYKLOQLESASKIOVKSIAQITAIIRKLEK
LKQELTASETDGPASEVFCNTLKLDIFISIAETEMATVLSLVSVLFRKAHEENVKQADL

CDS

tRNA

CDS

CDS

CDS

join(23327..23486,23729..23834,23917..24050,24190..24245,
24360..24446,24519..24839,25117..25236,25524..25658,
25789..25865,25960..26035,26177..26231,26364..26468,
26640..26809,26913..27015,27122..27234,27523..27724,
27894..28025,28107..28221,28349..28434,28504..28599,
28759..28895,28993..29095,29179..29258,29359..29437,
29528..29585,29895..29992,30151..30245,30343..30400)
/note="gene_id:MBK20.10"
/codon_start=1
/evidence=not_experimental
/product="SMC-like protein"
/protein_id="BAB11444.1"
Query Match 0.98; Score 24; DB 12; Length 78172;
Best Local Similarity 100.0%; Pred. No. 0.35; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;
Qy 1513 ctccaccacccacccgcctctc 1536
|||||
Db 63436 CTCACCCCCACCCGCGCTCCTC 63413

RESULT 13
AC073794_1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC073794 Accession AC073794
Fragment Name Begin End
AC073794_0 1 110000
AC073794_1 100001 210000
AC073794_2 200001 310000
AC073794_3 300001 407732
Continuation (2 of 4) of AC073794 from base 100001 (AC073794 Mus musculus clone RP23-

Query Match 0.98; Score 24; DB 75; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1599 tctgggctacatcccaactccatct 1622
|||||
Db 72300 TCTGGGCTACATCCCACTCCATCT 72277

RESULT 14
AC073822/c
LOCUS AC073822 247196 bp DNA HTG 18-JUL-2000
DEFINITION Mus musculus clone RP23-85K13, WORKING DRAFT SEQUENCE, 16 ordered
pieces.
ACCESSION AC073822
VERSION AC073822.2 GI:9256816
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 247196)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 247196)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 18, 2000 this sequence version replaced gi:8810439.
-----Genome Center

Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1769765
Center clone name: RPCI-23_85K13

Summary Statistics

Consensus quality: 236727 bases at least Q40
Consensus quality: 243596 bases at least Q30
Consensus quality: 244906 bases at least Q20
Estimated insert size: 243000; agarose-fp estimation
Estimated insert size: 246496; sum-of-contigs estimation
Quality coverage: 7.76 in Q20 bases; agarose-fp estimation
Quality coverage: 7.65 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 18947: contig of 18947 bp in length
* 18948 19047: gap of unknown length
* 19048 20569: contig of 1522 bp in length
* 20570 20669: gap of unknown length
* 20670 27716: contig of 7047 bp in length
* 27717 27816: gap of unknown length
* 27817 29329: contig of 1513 bp in length
* 29330 29429: gap of unknown length
* 29430 62474: contig of 33045 bp in length
* 62475 62574: gap of unknown length
* 62575 66039: contig of 3465 bp in length
* 66040 66139: gap of unknown length
* 66140 91228: contig of 25089 bp in length
* 91229 91329: gap of unknown length
* 91329 115460: contig of 24132 bp in length
* 115461 115560: gap of unknown length
* 115561 132567: contig of 17007 bp in length
* 132568 132667: gap of unknown length
* 132668 136189: contig of 3522 bp in length
* 136190 136289: gap of unknown length
* 136290 178101: contig of 41812 bp in length
* 178102 178201: gap of unknown length
* 178202 187292: contig of 9091 bp in length
* 187293 187392: gap of unknown length
* 187393 188720: contig of 1328 bp in length
* 188721 188820: gap of unknown length
* 188821 193055: contig of 4235 bp in length
* 193056 193155: gap of unknown length
* 193156 205419: contig of 12263 bp in length
* 205419 205519: gap of unknown length
* 205519 247196: contig of 41678 bp in length.

FEATURES

source
1..247196
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-85K13"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 56420 a 65394 c 66141 g 57737 t 1504 others
ORIGIN

Query Match 0.9%; Score 24; DB 75; Length 247196;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1599 tctgggggtacatcccaactcatct 1622

Db 243129 TCTGGGCTACATCCCACTCATCT 243106

RESULT 15

CNS0707X
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..993

/organism="Kluyveromyces marxianus"

/strain="CBS 712"

/varlety="marxianus"

/db_xref="taxon:4911"

/clone="AZ0AA007E05"

/clone_lib="AZ0AA"

/notes="end : T3"

BASE COUNT 266 a 378 c 155 g 190 t 4 others

ORIGIN

Query Match 0.9%; Score 23; DB 53; Length 993;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1074 ccacctccacctcccaatgca 1096

Db 328 CCACCTCCACCTCCACCAATGCA 350

Search completed: June 20, 2001, 23:38:51

CNS0707X 993 bp DNA STS 11-JAN-2001
T3 end of clone AZ0AA007E05 of library AZ0AA from strain CBS 712 of
Kluyveromyces marxianus, sequence tagged site.

AL423283

AL423283.1 GI:12206477

STS.

Kluyveromyces marxianus.

Kluyveromyces marxianus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 993)

Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.

and Dujon,B.

Genomic Exploration of the Hemiascomycetous Yeasts: 12.

Kluyveromyces marxianus var. marxianus

FEBS Lett. 487 (1), 71-75 (2000)

11152887

2 (bases 1 to 993)

Soucliet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,

Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,

de-Montigny,J., Dujon,B., Durrens,P., Lepingie,A., Llorente,B.,

Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,

Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

Wincker,P. and Weissenbach,J.

Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of

yeast species for molecular evolution studies(1)

FEBS Lett. 487 (1), 3-12 (2000)

11152876

3 (bases 1 to 993)

Genoscope.

Direct Submission

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This STS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

Location/Qualifiers

1..993

/organism="Kluyveromyces marxianus"

/strain="CBS 712"

/varlety="marxianus"

/db_xref="taxon:4911"

/clone="AZ0AA007E05"

/clone_lib="AZ0AA"

/notes="end : T3"

BASE COUNT 266 a 378 c 155 g 190 t 4 others

ORIGIN

Job time: 19119 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 15:04:28 ; Search time 3470.44 Seconds
(without alignments)
11699.624 Million cell updates/sec

Title: US-09-425-501-1
Perfect score: 2625
Sequence: 1 cttctctgcacttcgcat.....aatggttcattttaaaagt 2625

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_rod:*
33: em_hum1:*
34: em_hum2:*
35: em_hum3:*
36: em_hum4:*
37: em_hum5:*
38: em_hum6:*
39: em_hum7:*
40: em_in:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_v12:*
60: gb_vtg1:*
61: gb_vtg2:*
62: gb_vtg3:*
63: gb_vtg4:*
64: gb_vtg5:*
65: gb_vtg6:*
66: gb_vtg7:*
67: gb_vtg8:*
68: gb_vtg9:*
69: gb_vtg10:*
70: gb_vtg11:*
71: gb_vtg12:*
72: gb_vtg13:*
73: gb_vtg14:*
74: gb_vtg15:*
75: gb_vtg16:*
76: gb_vtg17:*
77: gb_vtg18:*
78: gb_vtg19:*
79: gb_vtg20:*
80: gb_vtg21:*
81: gb_vtg22:*
82: gb_vtg23:*
83: gb_vtg24:*
84: gb_vtg25:*
85: gb_vtg26:*
86: gb_vtg27:*
87: gb_vtg28:*
88: gb_vtg29:*
89: gb_vtg30:*
90: gb_vtg31:*
91: gb_vtg32:*
92: gb_vtg33:*
93: gb_vtg34:*
94: gb_vtg35:*
95: gb_vtg36:*
96: gb_vtg37:*
97: gb_vtg38:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2625	100.0	2625	91	D87459
2	1782.6	67.9	2469	94	AF290877
3	1758	67.0	1758	89	AF134303
C 4	861.4	32.8	151580	82	AL590009
C 5	861.4	32.8	192017	82	AL590549
6	837.4	31.9	213475	87	AC019100
7	751.2	28.6	143812	66	AC021847
8	282.6	10.8	1509	85	AB026543

Db 541 CTTTCGAGGTTCTACAATTCAGACGAGCAGCTTTTCGATCGCAAGACTTTCGCCATTTC 600
QY 601 cattagagagacgtacgtatgttgaaacagcctccacccctcaataactactcactcctt 660
Db 601 CATTACAGAGACGTACGATGTTGTGACAGCCTCCACCTCTCAATATACACTCCCTT 660
QY 661 atagagatgatgtaagaaggtctgaagttttataccaactccttctgtattttttgatc 720
Db 661 ATAGATGATGTGTAAGAAGGTCGTAAGTTTATACCAATCCTCTGCTATTTCCTTGATC 720
QY 721 tatgaaagaaataatgttgcagatacagagataagaggaagaaagagaaacagaga 780
Db 721 TATGGAAGAAATAATGTTGCAGATACAGAGGATAGAGGAAGGAAAGAGGAGCAGA 780
QY 781 agcagaaaaatctagatcgctcatgaaccagaaaaagtgccaaagagacccatgaca 840
Db 781 AGCAGAAAAATCTAGATCGTCTCATGACCAACAGAAAAAGTGCCAGAGCACCTCATGACA 840
QY 841 ggcggcgaagatggcgaagcgtggcccaaggtccagagctggctgaagatgatgctaactc 900
Db 841 GCGCGGAGAATGGCAGAGCTGGCCCAAGTCCAGAGCTGGCTGAAGATGATGCTTAATC 900
QY 901 tcttataagcatattgaagttgctaatggcccgagcctctctatttgaacaagacctc 960
Db 901 TCTTACATAGCATATTGAAGTTGCTTAATGGCCCGAGCCTCTCAITTTGAAACAGAGCCTC 960
QY 961 agacatacgtggatcatatgatggatcttactcaactttctgcttgcctatttagtcaga 1020
Db 961 AGACATACGTGGATCATATGATGGATCTTACTCACCTTTCGCTTGCCTTAGTCAGA 1020
QY 1021 tgagtgagctctgactagagctgaggaaggttatagtcagacacatgaaccacctc 1080
Db 1021 TGAGTGAGCTTCTGACTAGAGCTGAGGAAGGGTATTAGTCAGACACACATGAACCACTC 1080
QY 1081 cactccaccaatgcattgagagagagatgcaaaacggatgataccaccctgtatcagttctg 1140
Db 1081 CACTCCACCAATGCATGAGAGAGAGATGCAAAACGGATACCCACCTGTATCAGTTCTG 1140
QY 1141 ctacaggtttgataaaaaatcgccctcagtcaccagctacagcagacaacacctggttg 1200
Db 1141 CTACAGCTTTGATAGAAAATCGCCCTCAGTCACCAGCTACAGCAGAACACCTGTGTTG 1200
QY 1201 tgagcccaactccccacactctccaccactcttccactcttccactctgcttgcacttccat 1260
Db 1201 TGAGCCCACTCCGCCACCTCTCCACCACTCTTCCATCTGCTTGTCAACTTCCTCAT 1260
QY 1261 taagagcttcaatgacttcaactctccctccctccagtaacctcccccaactccacctccag 1320
Db 1261 TAAGAGCTTCAATGACTTCAACTCTCCCTCCAGTACCTCCCACTCCCACTCCAG 1320
QY 1321 ccactgctttgcaagctccagcaggtaccacacctccagctcctctcagattgcccctg 1380
Db 1321 CCAGTCTTTGCAAGCTCCAGAGTACCACCACTCCAGCTCCTCTTCAGATTGCCCTG 1380
QY 1381 gagtcttccaccagctcctctcccaattgcaacctctcttagtacagccctctccaccag 1440
Db 1381 GAGTTCTTCAACCCAGCTCTCTCCCAATTCACCTCCTCTAGTACAGCCCTCTCCACCA 1440
QY 1441 tagctagagctgcccagatgtgagactgtaccagttccactccaccccaagggtaag 1500
Db 1441 TAGCTAGAGCTGCCAGTATGTGAGACTGTACAGCTTACAGTTTCACTCCCAAGAGTGAAG 1500
QY 1501 ttccagggctgctccaccccccacccgctcctctgctccacactggattcgacct 1560
Db 1501 TTCAAGGGCTGCTCCACCCCACTCCCTCTCTGCTCCAGCTTGGCATTTCGACCAT 1560
QY 1561 catcacctgtcacagttcacagctcttctcatctcctcctctggctacatcccaactccat 1620
Db 1561 CATCACCTGTACAGTTTACAGCTTCTGCTCATCTCTCTCTGGCTTACATCCAACTCCAT 1620
QY 1621 ctactgcccaggtcccacatggttccatataatgctcccatctcctccatccacaagttatc 1680

Db 1621 CTACTGCCCGAGGTCCCCATGTTCCATTTAATGCCTCCATCTCCTCCATCAAGTTATAC 1680
QY 1681 ctgcttctgagccaaagcgccatcccatcaacctactgtaatactcagtgatgccagagtg 1740
Db 1681 CTGCTTCTGAGCCAAAGCGCCATCCATCAACCTACTGTGTAATCAGTGTATGCCAGGAGTG 1740
QY 1741 tgcactcgaagcaatacagaaaaaggtattcagctacgcaaaagtagaagagcagcgtgaac 1800
Db 1741 TGCTACTGGAAGCAATACGAAAAGGTATTACAGCTAGCAAGTAGAAGAGCAGCGTGAAC 1800
QY 1801 aggaagctaaagcatgaacattgaaacgagtgtgccaccatcctgctcgcgcgtatg 1860
Db 1801 AGGAAGCTTAAGCATGAACGATTTGAACAGATGTTGCCACCCTCTGCTGCCGCTATTG 1860
QY 1861 ctgttgaatatagtgattcgggaagatgattcagaatttgaatgaagtagattggttggagt 1920
Db 1861 CTGTTGAATATAGTATTCGGAAGATGATTACAGAAATTTGATGAAGTAGATTGCTGGAGT 1920
QY 1921 aagaaaaatgcattgataataattacaaaaactgaaatgcaaaaatgctcttctgtggtgtgt 1980
Db 1921 AAGAAAAATGCATTGATAAATATTACAAAACGCAAACTGAATGCAAAATGCTTGTGTGCTTGT 1980
QY 1981 tccctgaaaaatgtttggtcattcagttgttcttcttcttcttcttataataaataaccc 2040
Db 1981 TCCCTGAAAATGTTGGTCACTTCTAGTGTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2040
QY 2041 tttctccataaacttttgaatttcttaaggaataattagcatacatttcaaaactaaatgt 2100
Db 2041 TTTTCTCTCCATAACTTTTGATTTCTTAAGGAAAATATTAGCATACATTTTCAAACTAAATGT 2100
QY 2101 tttaagtggtgcttatottt 2160
Db 2101 TTTTACAGTGGCTTATCTTTTTCCTCCCTGAAAGAACTAAATTTGGTCAAAATAAACCACT 2160
QY 2161 aagtttaagcatggcagcgtgtttagagtagcagattcagtttttttqatatattctta 2220
Db 2161 AAGTATTAACTAGCATGGACAGCTGTTGTAGAGTAGCAGATTTCAGTTTTTGTGATATCTTA 2220
QY 2221 attgtgacttttgaaattttaaaagaaagcaactgaaatgaaatccttgagggca 2280
Db 2221 ATTGTGTACTTTTGTGAATTTTAATTTAAAGAAAGCAACTGAAATTTGAAATCTTGAGGGCA 2280
QY 2281 gctgctactactaatgagccttattccacttctctgattgttttaaaagaaagaaacactgccc 2340
Db 2281 GCTGTATCTACTAATGAGCCTTATTCATTTCTCTGATGTTTAAAGAAAGAAACACTGCC 2340
QY 2341 ttgaattacgaatacactcagaaagtacattagcttctgtagttgaaattctcttaaaag 2400
Db 2341 TTGATTATAGAAATACACTCAGAAAGTACATTTAGCTTTGTAGTGTGAATTTCTTTAAAG 2400
QY 2401 gaatgcttgaatttttctcattattgttttttttttttttttttttttttttttttttttttt 2460
Db 2401 GAATGCTTGAATTTTTCATTTATTTTATTTTATTTTATTTATTTATTTATTTATTTG 2460
QY 2461 tttagcagtatccccctccccacttatatatgtgtgatataattgttgccttgcctatagga 2520
Db 2461 TTTTAGCAGTATCCCTTCCCACTTATATTTGTGTATATGATTTGCTTGTGCTCTATAGGA 2520
QY 2521 gttaaaaacttttccatgtgaaataactctgaacttaaacatacatgaacttacataactg 2580
Db 2521 GTTAAAAACCTTTTCCATGTGAAAATCTCTGACTTAAACATACATGTAACTTACATAACTG 2580
QY 2581 ttaagaataacagcttgatttaataaaatggttcatattttaaagt 2625
Db 2581 TTAAGAATAACAGCTCTGATTTAATAAATGGTTCTATTTTAAAGTT 2625

RESULT 2

AF290877 LOCUS AF290877 2469 bp mRNA ROD 10-OCT-2000
DEFINITION Mus musculus WAVE-1 mRNA, complete cds.
ACCESSION AF290877
VERSION AF290877.1 GI:9931545

KEYWORDS

SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 2469)

AUTHORS

Westphal, R.S., Soderling, S.H., Alto, N.M., Langeberg, L.K. and

Scott, J.D.

TITLE

Scar/WAVE-1, a wiskott-aldrich syndrome protein, assembles an

actin-associated multi-kinase scaffold

JOURNAL

EMBO J. 19 (17), 4589-4600 (2000)

MEDLINE

20428428

PUBMED

10970852

REFERENCE

2 (bases 1 to 2469)

AUTHORS

Soderling, S.H., Westphal, R.S. and Scott, J.D.

TITLE

Direct Submission

JOURNAL

Submitted (27-JUL-2000) Howard Hughes Medical Institute/Vollum

Institute, Oregon Health Sciences University, 3181 Sam Jackson Park

Rd., Portland, OR 97201, USA

FEATURES

Source

1..2469

/organism="Mus musculus"

/db_xref="taxon:10090"

CDS

1..1680

/note="Kinase-anchoring protein; Abl-binding protein,

Arp2/3 complex activator"

2

/codon_start=1

/product="WAVE-1"

/protein_id="MAG02214.1"

/db_xref="gi:9931546"

/translation="MPLVKNRIDPRHLCHTALPRGKNELECVTNISLANIROLSSL

SKYADIFGELFNASHFSFRVNSLOERDRLSVTOLDPEBELSLQDITMKAFR

STIQDQOLFPLPI PLOETVCEOPPLNITLTPYRDDGKEGLFYTNPSYFFDL

WKEWLODEKREKOKNDLDRHEPEKVPRAHDRREKQALQGPDLAEDDA

DLLHKHIZVANGPASHETREQTVYDMDGYSLSALPFSOMSELLRAERVLVRPH

EPPPPMHMGADKAPPTCISATGLIENRQSPAGRTVFVSPPTPPPPPLPSA

.LSTSLSRASMTSTPPPPPPPPPPATLQAPVPPPIQIAPGLVHPAPPIAPP

LVQSPVPAARAACVETVPVHPLQGEVQGLPPPPPLPPPPRSPSPVAVAAAH

PPSGPLHAPAPHPMPSPSPSVLPASEPKRHPSTLPVTSIDRSPVLEAIRKG

IQLRVEPQREQAKHERIENDVATILSRRTAVEYSODESDSEFDEVDWLE"

BASE COUNT 659 a 638 c 521 g 650 t 1 others

ORIGIN

Query Match 67.9%; Score 1782.6; DB 94; Length 2469;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 2120; Conservative 1; Mismatches 250; Indels 88; Gaps 10;

QY 243 atgcgcgtagtgaagaacacgcagctatcctagcactgtgcccacacagcactgcctaga 302
Db 1 ATGCCGTTGGTGAAGAAGAACATCGACCTCGGCACCTGTGCGCACACAGCAGTCCAGG 60

QY 303 ggcattaagaatgaactggaatgttaaccaataattctcttggaataataattagacaa 362
Db 61 GGCATTAGAAGCAACTGGAATGTGTAAACCAACATTTCCCTGGCAATATAATTAGACAA 120

QY 363 ctaagttagcctaagtaaatatctgaagatatatttgagaattattcaatgaagcacat 422
Db 121 CTGAGTAGCTTAAGTAGTATGCTGAAGACATATTGAGAAATATTCAATGAAGCACAC 180

QY 423 agttttcttcagagtcacactcatctgaagaacgtgtgacgcgtttatctgttagtgtt 482
Db 181 AGTTTTTCCTTCAGAGTTAACTCGTTGGAAGAGCGGTGGACCGATTGCTGTTAGTGTG 240

QY 483 acacagcttgatccaaagaagaagaattgtcttttgcaagatatatacaatgaggaagct 542
Db 241 ACACAGCTTGATCCTAAGGAGGAAGAACTGTCTTTGCAAGATATAACGATGAGAAGGCT 300

QY 543 ttccgaagtcttacaattacaagaccagcagcttttccgacgcaagactttgctattcca 602
Db 301 TTCCGGAGCTCCACAAATTCAGATCAGCAGCTTTTCAGCCGCAAGACGTTGCTATTCCA 360

QY 603 ttacaggagaactgactgtgtttgtgaacagcctccactcactcactcactcacttat 662

Db 361 TTGCAAGAAACATACGATCTGTGTGAGCAACCTCGCCTCTCAATATCTCACCCCTTAC 420

QY 663 agagatgatgttaagaaggtctgaagttttatataccaatcctctgattttttatctata 722

Db 421 AGGGATGATGTGTAAGGAAGGCTGAAGTTTATACCAATCTCTTCATATTTCTTTGATCTG 480

QY 723 tgaagaagaaaaattgtcaagatacacagagataagagaagaaaaagagagagagag 782

Db 481 TGGAAAGAAACATGTTTCAAGATACAGAGGACAGAGGAGGAGGAGGAGGAGGAGGAG 540

QY 783 cagaaaaatctagatcgtcctctcatgaaccagaaaaagtgccaagagcctctcatgacagg 842

Db 541 CAGAAAAATCTAGACCGCTCTCATCAACAGAGAAAGTGCCAAGAGCAGCTCATGACAGG 600

QY 843 cggcagagaatggcagaagctggcccaaggtccagagctggctgaagatgagtgctaatctc 902

Db 601 CGCGGGGAATGCGCAAGCTGGCCCAAGTCCAGAGCTGGCTGAGGATGACGCTGACCTC 660

QY 903 ttacataagcatattgaagtgtctaattgcccagcctctcattttgaaacaagacacctcag 962

Db 661 CTACACAAGCATATTGAAGTTGCCAATGCCAGGCTCTCATATGACACAAGGCCACAG 720

QY 963 acatacgtggatcatatggatggatcttactcaactttctgcttgccttgcctttagtcagatg 1022

Db 721 ACATACGTTGGATCATATGACGAGTCTGCTACTCTCTGCTTGCCTTGCCTTGCCTCAGATG 780

QY 1023 agtgagcttctgactagagctgaggaaggggtatttagtcagacacacatgaaccaacctcca 1082

Db 781 AGTGAACCTTCTGACAGAGCTGAGGAGAGGGTCTTAGTCAGACCCCATGAACCTCTCTCG 840

QY 1083 cctccaccaatctgagcagcagagatgcaaaacagatcaccacactgtatcagttctctgt 1142

Db 841 CCTCCACCAATGATGAGGAGGAGATGCAAAAGGCCACACACCTGTATCAGTTCTGCT 900

QY 1143 acaggtttgatagaaaaatgcgcctcagtcacacagctacagcagcagcagcagcagcagc 1202

Db 901 ACAGGTTTGTATAGAAATCGGCCAGTCCCGCAGTCCCGCAGTGCAGGAGAGACACCTGTGTTGTG 960

QY 1203 agccccactccccactccccactccccactccccactccccactccccactccccactcatta 1262

Db 961 AGCCCCACTCCCCACCTCTCTCCCCACCTCTCTCGCTCTGCTTGTCTTCTTCTTCTTCACTA 1020

QY 1263 agagcttcaatgaactcaactctctccctccctcagtcactccccactccccactccccagcc 1322

Db 1021 AGAGCTTCAATGACCTCAACTCTCTCCCCACAGTCCCTCTCTCGGCCCTCTCTCCCCAGCC 1080

QY 1323 actgcttgcagctccagcagctaccacacacacacacacacacacacacacacacacacac 1382

Db 1081 ACTGCTTTTGCAGGCCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 1140

QY 1383 gttcttcaaccagctcctctcccaattgacactcctctagtcagacacacacacacacacac 1442

Db 1141 GTTCTTCAACCAGCTCTCTCTCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200

QY 1443 gctagagctgccccagctgtgagagctgtacagttcactccactccccacacagtgaaagt 1502

Db 1201 GCTCGAGCTGCCAGATGCGAGACTGTGCGAGTTCACCCACATCCCGCAAGGTGAAGTG 1260

QY 1503 caggggctgcctccacccccac 1562

Db 1261 CAGGGGCTGCTCCACCCCGCCACACCGCTCTCTTACCTCGGCTTGGCATCGCGGCATCA 1320

QY 1563 tcactgtcacagttacagctctgtctatcctcctcctcctcctcctcctcctcctcctcctcct 1622

Db 1321 TCTCTGTGCGGAGTTCAGCTCTTGTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380

QY 1623 actgccccaggtctcccatgttccatttaagtccctcctcctcctcctcctcctcctcctcctc 1682

Db 1381 ACTGCCCGGGTCCCATGCTCCATTAATGCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440

QY 1683 gcttctgagcacaagcgcacatccatccacacctacactgtaactgagtgagcagagagtg 1742

Db 1441 GCCTCTGAGCCAAAGCGCCATCCATCCACCCCTGCTGTATATCAGTGACGCAAGGAGTGTG 1500

QY 1743 ctactggaagcaatacgaagaaagtattcagctacgcgaagtagaagcagcagcgtgaacag 1802

Db 1501 CTCTCTGGAAGCCATCCGAAAGGATATTCAGCTTCCGAAAGTGAAGAGCAGCGTGAACAG 1560

QY 1803 gaagctaaagcagcaagcattgaaacagatgttgccaccatcctgtctgcgcgtattgct 1862

Db 1561 GAAGCAAAACACAGCGCGGATGAAGACGAGCTGGCCACCATCTGCTCGCGCATGTCT 1620

QY 1863 gttgaataatagttgattcgcgaagatgattcagaattgtagaagtagattggttggaagtaa 1922

Db 1621 GTGGACTATAGCACTCGGAAGATGACTCGGAGTTCGATGAAGTACGAGTGGCTGGAGTAA 1680

QY 1923 gaa----- 1925

Db 1681 GAAAGCCACAGTGACTTGGAAAGATGACTCGGAGTTCGATGAAGTACGAGTGGATGGAGTAA 1740

QY 1926 ---aaatgcattgataaataattacaaaactgaatgcaaatgtctcttgggtgctgttc 1982

Db 1741 GACAGCGCGCTTGCTCTGTATTGCAAGGCTGAATGCACATGTCCTCGTGGCGCTTGTTC 1800

QY 1983 cttgaaaatgtttgctcattcagttgattgcttcttcttcttataataaataagaccctt 2042

Db 1801 TTTGAAATGTTGGTTCATCTAGTCTTGTCTCTCTATCCATCAATGAATAACCCCT 1860

QY 2043 ttctccataaacttttgattcttaagaaaataattagcatataattcacaataaattgttt 2102

Db 1861 TTCTCTCCATTATTTTGTATTTATGAAAGTATTAGCATATATTTCAAACATAAATTT 1920

QY 2103 tacagtgttattctt---ttttcccccctgaaaagactaatttggtgcaataaaccact 2160

Db 1921 T---GTGGCTTACTTTAAATTTTCCCCCTGGAAAAATAAATTTGGTCCAAATAAATCT 1977

QY 2161 aagtataagcatggagcagctgttgtagagtagcagattcagtttttggatatactta 2220

Db 1978 AAGTATTAGCATGGAAGCCGTTATTAGTAGAGCATTCAG-TTTTGTATATTCCTTA 2036

QY 2221 attgtactttgtgaattt-----aatttaaagaagcaactgaaattg 2266

Db 2037 ATGTGCACTTTCTGTATTTTCTTTTAAATTTTGAAGACGCAACTGAGACTG 2096

QY 2267 aaattcgtgagcagctgtatctactaataagccttattccatttctcctgatgttttaaaa 2326

Db 2097 GAATCTTGAAGGAGCTGTCTCTGGAAACGAGCTTATTCGCTTCTTGTATGTTTAAAT 2156

QY 2327 gaagaacactgccttgattatcagaataacactcagaagacattatagctttagttgtt 2386

Db 2157 GAAGAAACACTGCTCGATTATATGAAATACATCAGAAAGAAACATGTAGCTTGTAGTGT 2216

QY 2387 gaattctctaaaggaatgcttgaatttttttctatttattgttttttataactt 2446

Db 2217 GAATCTCTTAAGGAATGCTT---ACCTTGTGTATCATTTTGTGTTTGAATACTT 2273

QY 2447 gccttatttgaaattgtagcagctatcccttccacttatatttattgtagtatgtttt 2506

Db 2274 GCCTTATTTGACGTTTAGCAG-AACCCCTTCCACTTATATATTGTTGCTACAGTTT 2332

QY 2507 gcttgctatagagagtaaaacttttccatttccatgtgaataactctgaacttaaaacatactg 2566

Db 2333 GCTTGCTATAGAAGTT-TAAATTTTCCATGTGAATACTCAGA-ATAAGCATATGTGT 2390

QY 2567 aacttacataactgtttaagaataacagctgtgatttaataaattgtttcatttttaaaagt 2625

Db 2391 GACTTACATAA--TGTGAGAAATTACAATCTGATGTAATAAAGCGTTCAITTTAAAGCT 2447

RESULT 3

LOCUS AF134303 1758 bp mRNA PRI 14-JUN-1999

DEFINITION Homo sapiens Scarl (K1AA0269) mRNA, complete cds.

ACCESSION AF134303

VERSION AF134303.1 GI:4927209

human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1758)

AUTHORS Bear, J.E., Rawls, J.F. and Saxe, C.L. III.

TITLE SCAR, a WASP-related protein, isolated as a suppressor of receptor defects in late Dictyostellium development

JOURNAL J. Cell Biol. 142 (5), 1325-1335 (1998)

MEDLINE 98402540

PUBMED 9732292

REFERENCE 2 (bases 1 to 1758)

AUTHORS Machesky, L.M. and Insall, R.H.

TITLE Scarl and the related Wiskott-Aldrich syndrome protein, WASP, regulate the actin cytoskeleton through the Arp2/3 complex

JOURNAL Curr. Biol. 8 (25), 1347-1356 (1998)

MEDLINE 99108169

PUBMED 9889097

REFERENCE 3 (bases 1 to 1758)

AUTHORS Machesky, L.M. and Insall, R.H.

TITLE Direct Submission

JOURNAL Submitted (12-MAR-1999) Biochemistry, University of Birmingham, Birmingham B15 2TT, England

FEATURES

Source

1. .1758

gene /organism="Homo sapiens"

1. .1758 /db_xref="taxon:9606"

gene /gene="K1AA0269"

CDS 1. .1680

/gene="K1AA0269"

/function="activates dendritic nucleation of actin filaments by the Arp2/3 complex"

/note="WASP-family protein; WAVE; similar to Dictyostellium suppressor of cyclic-AMP receptor"

/codon_start=1

/product="Scarl"

/protein_id="AAD33052.1"

/db_xref="GI:4927210"

/translation="MPLVKRNDPRHLCHTALPRGINKNECVTNISLANIIRLSSLL SKYAEIDFGELENEAHSFRVNSLOERVDRLSVSVTQDLPKEELSLODIPMKAFR STIQDQQLDFDKTLPIPLQETDYCEQPPNLTTPYRDDKRGKLFKTFNYSYFDL WKEMLQDTEDEKREKROKOKNDPHEKVPKPRAPHRDRKWKGLAQGLAEEDDA NLLKHIEVANGPASHFETPQYVDHMDGYSLSALPFSQMSSELLTRAEEVLVRPH BPPPPPMHAGDAPKIPITCISATGLIENRQSPATGTPVEVSPTPPPPPPLPSA LSTSSIRASMTSTPPPPPPPPPPATLQAPVPPPPAPLOIAPCVLHPAPPTAPP LVQSPPPVAAAPVCEVTVVHPLPQGEVGLPPPPPPPPPPPPPPPPGIRPSPFVITAAH PPSPPTPTSTAPGPHVLPMPSPSPSPQVIPASEPKRHPSTLPIVSDARSVLLEAIRKG IQLRKVEQREQEAKHERIENDVATILSRRIAVEYSDSEDFEVDWLE"

BASE COUNT 506 a 467 c 343 g 442 t

ORIGIN

Query Match 67.0%; Score 1758; DB 89; Length 1758;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 atgcccctagtgaaagaaacatcgatccttagcacttggtccacacagcagcctcaga 302

Db 1 ATGCCGCTAGTGAAGAAAGAAACATCGATCCTAGGACTTGTGCCACACAGCAGCTGCCTAGA 60

QY 303 ggcattaaagaatgaactggaatgtgaaccaatatttcccttggaataataaattagacaa 362

Db 61 GGCATTAAAGAACTCACTGGAAATGTGAACCAATATTTCCTTGGCAATAATAATTAGACAA 120

QY 363 ctaagtagcctaagtaataatgctggaagatatatttggagaattattcgaatgaagcaat 422

Db 121 CTAAGTAGCCTAAGTAATAATGCTGAAGATATATTGGAGAATTATTCAATGAAGCACAT 180

QY 423 agtttttcttcagagcactcaactcattgcaagaacgctgtggaccgtttattctgttagtt 482

Db 181 AGTTTTTCTTTCAGAGTCAACTCATTTCCAAAGACGCTGTGACCGTTTATCTGTAGTGT 240

```
QY 483 acacagcttgatccaaagaaagaattgtcttttgcaagatataaacaatgagaaagt 542
|||||
Db 241 ACACAGCTTGATCCAAAGGAAGAAATGTCTTTGCCAGATATACAAATGAGGAAGCT 300
|||||
QY 543 ttccgaagtctacaaatcaaacaccagcagcttttcgatcgcaagactttgctattcca 602
|||||
Db 301 TTCCGAAGTCTTACAATTCAGACCAGCAGCTTTTCGATCGCAAGACTTTTGCTATTCCA 360
|||||
QY 603 ttacaggagacgtacagatgttttgtaaacagcctcaacctctcaataatactactccttat 662
|||||
Db 361 TTACAGGAGACGTACGATGTTTGTGAACAGCCTCCACCTCTCAATATACTACTCTCTTAT 420
|||||
QY 663 agagatgatggtaaaagagctgtaagtctttataccaatccctcgattctttgatctca 722
|||||
Db 421 ACAGATGATGGTAAAGAGGCTCGAAGTTTATACCAATCCTTCGTATTTCTTTGATCTA 480
|||||
QY 723 tggaaagaaaaatgttgaagatacacagagataagaggaagaaagaggagcagaag 782
|||||
Db 481 TGGAAAGAAAAAATGTTGAAGATACAGAGGATAAGAGGAAGGAAAGAGGAGCAGAA 540
|||||
QY 783 cagaataatctagatcgctcctcatgaaccagaaaaagtgcgaagcaccctcatgacagg 842
|||||
Db 541 CAGAAAAATCTAGATCGTCTCATGAACACAGAAAAAGTGCCAAAGACACCTCATCACAG 600
|||||
QY 843 cggcgagaatggcagaagctggcccaaggtccagagctggtggaagatgagtctaatctc 902
|||||
Db 601 CGCGGAGAATGCAGANCTGGCCCCAGGTCCAGAGCTGGCTGAAGATGATGCTTAATCTC 660
|||||
QY 903 ttacataagcatattgaagtgtctaattggccagcctctcattttgaacaagaacctcag 962
|||||
Db 661 TTACATAAGCATATTGAAGTTGCTAATGCCAGCCTCTCAATTTGAAACAAGACCTCAG 720
|||||
QY 963 acatcgtgatacatagatgagcttactcaactcttctgcttcccttgcttagtcagatg 1022
|||||
Db 721 ACATAGTGATCATATGATGGATGATCTTACTCACCTTTCTGCTTGCCTTTAGTCAGATG 780
|||||
QY 1023 agtgaagctctgactagagctgaggaagggattagtcagaccacatgaaccactcca 1082
|||||
Db 781 AGTGAGCTTCTGACTAGAGCTCAGGAAGGGTATTAGTCAGACCACATGAACCACTCCA 840
|||||
QY 1083 cctccaccaatgcattggagcaggagatgcacaaacagatacccaacctgtatcagttctgt 1142
|||||
Db 841 CTTCCACCAATGCATGGAGCAGAGATGCAAAACCGATACCCACCTGTATCATGTTCTGT 900
|||||
QY 1143 acaggtttgatagaaaatcgccctcagtcaccagctacagcagacacacctgtttgtg 1202
|||||
Db 901 ACAGGTTTGATAGAAAATCGCCCTCAGTCACCAGCTACAGGCGAAGACCTGTGTTTGTG 960
|||||
QY 1203 agccccactccccacactctctccacacactcttccatctgcttgcacttctcactta 1262
|||||
Db 961 AGCCCCACTTCCCCCACCCTCTCCACCACTCTTCCATCTGCCCTTGCTCAACTTCTCATTA 1020
|||||
QY 1263 agagcttaactgaacttaactcctccccctccagtacctccccacactccacactcagcc 1322
|||||
Db 1021 AGAGCTTCAATGACTTCAACTCTCTCCCTCCAGTACCTCCGCCACCTCCACCTCCAGCC 1080
|||||
QY 1323 actgcttgcaagctccagcagatcacacacactccagctcctcttcagattgacctctgga 1382
|||||
Db 1081 ACTGCTTTGCAAGCTCCAGCAGTACACACCTCCAGCTCTCTTCAGATTGCCCTTGA 1140
|||||
QY 1383 gttcttcaacccagctcctcctccaaatggcacctcctctagtacagccctctccaccagta 1442
|||||
Db 1141 GTTCTTCAACCCAGCTCTCTCTCAATTCACCTCTCTAGTACAGCCCTCTCCACCAGTA 1200
|||||
QY 1443 gctagagctgccccagtgatgagagctgtaccagttcatccaactcccccaaggtgaagtt 1502
|||||
Db 1201 GCTAGAGCTGCCCCAGTATGTGAGAGTGTACAGTTTATCCACTCTCCACAGAGTGAAGTT 1260
|||||
QY 1503 cagggtctgctcccccacccacacacgctcctctgctccacactgagcttcgacatca 1562
|||||
Db 1261 CAGGGGCTGCTCTCCACCCACCCACCGCTCTCTGCTTCCACTTGGCATTCGACCATCA 1320
|||||
```

```
QY 1563 tcacgtgcacagttacagctcttctcatcctcctctgggctacatcccaactccatct 1622
|||||
Db 1321 TCACCTGTACAGTTTACAGCTTCTTCTCATCTCCTCTGGGCTACATCCAACCTCCATCT 1380
|||||
QY 1623 actgccccaggtcccccatttccattaatgctcctcattctcctccatccaaagtataact 1682
|||||
Db 1381 ACTGCCCCAGGTCCCATGTTCCATTAAATGCTCCATCTCTCCATCACAAAGTTATACCT 1440
|||||
QY 1683 gcttctgacgaacacgcacatccatcaacctacatgaatcaatcagtgatgcaggaagtgtg 1742
|||||
Db 1441 GCTTCTGAGCCAAAGCCCATCCATCAACCCCTACCTGTAATCAGTGATGCCAGGAGTGTG 1500
|||||
QY 1743 ctactggaagcaatacgaaggtatttcagctacgcaagtagaagagcagcagctgaacag 1802
|||||
Db 1501 CTACTGGAAGCAATACGAAAGGTATTTCAGCTACGCCAAAGTAGAAGAGCAGCGTGAACAG 1560
|||||
QY 1803 gaagctaagcatgaacgattgaaacgattgttgccacatcctctctcgcgattgct 1862
|||||
Db 1561 GAAGCTAAGCATGAACGCAATTGAAACACGATGTTGCCACCATCTCTCGCGGTATTGCT 1620
|||||
QY 1863 gttgaataatgatttcggaagatgattcagaatttcagaatttgatgaagttagttggtgagtaa 1922
|||||
Db 1621 GTTGAATATAGTGTATTCGGAAGATGATTTCAGAAATTGATGAAGTAGATGTTGTTGAGTAA 1680
|||||
QY 1923 gaaaaatcattgataataattacaaaactgaatgcaaatgtccctttgtggtgctgttc 1982
|||||
Db 1681 GAAAAATCATTTGATAAATATTACAAAACTGAATGCAAAATGTCTTTGTGTTGCTTGTTC 1740
|||||
QY 1983 cttgaaaaatgtttgtgta 2000
|||||
Db 1741 CTTGAAAAATGTTTGTGTCA 1758
|||||
```

RESULT 4
AL590009/c

LOCUS Homo sapiens chromosome 6 clone RP11-181P4, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.

DEFINITION

ACCESSION AL590009

VERSION AL590009.4 GI:13568304

KEYWORDS HTGS; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 151580)

AUTHORS Corby,N.

TITLE Direct Submission

JOURNAL Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerquest@sanger.ac.uk

COMMENT On Apr 9, 2001 this sequence version replaced gi:13446548.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: ba181P4

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 149544 bases at least Q40

Consensus quality: 150676 bases at least Q30

Consensus quality: 150962 bases at least Q20

Insert size: 151080; sum-of-contigs

Insert size: 148237; agarose-fp

Quality coverage: 5.89x in Q20 bases; sum-of-contigs Quality coverage: 6.00x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces


```

DEFINITION Homo sapiens BAC clone RP11-443K8 from 2, complete sequence.
ACCESSION AC019100
VERSION AC019100.4 GI:9454638
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 213475)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 213475)
AUTHORS Kalicki, J., Drone, K. and Belter, E.
TITLE The sequence of Homo sapiens BAC clone RP11-443K8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 213475)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 213475)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 213475)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 213475)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 26, 2000 this sequence version replaced gi:7631000.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0443K08
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The RP11-443K8 BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E.,
Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
SOURCE INFORMATION:
The RP11-443K8 BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E.,
Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

```

approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-465011. Actual start of this clone is at base position 1 of RP11-443K8; actual end is at base position 213475 of RP11-443K8.

The clone RP11-443K8 contains a transposon which has been omitted from the submitted sequence. The transposon should insert after base position 88497.

FEATURES	Location/Qualifiers
source	1..213475
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
repeat_region	/clone="RP11-443K8"
	/clone_lib="RP11-443K8"
	1525..1702
	/rpt_family="MIR"
repeat_region	3044..3342
	/rpt_family="Alu"
repeat_region	3343..3459
	/rpt_family="MIR"
repeat_region	4351..4676
	/rpt_family="Alu"
repeat_region	5209..5380
	/rpt_family="MIR"
misc_feature	7249..7547
	/note="similar to EST AA411798 (NID:g2070413) zt67e02.s1"
misc_feature	7718..8054
	/note="similar to EST AA402120 (NID:g2056976) zt67e01.r1"
repeat_region	8373..8522
	/rpt_family="Alu"
repeat_region	8962..9242
	/rpt_family="MaLR"
repeat_region	9292..9531
	/rpt_family="MaLR"
repeat_region	9640..9962
	/rpt_family="MaLR"
repeat_region	9960..10087
	/rpt_family="MaLR"
repeat_region	10093..10544
	/rpt_family="MaLR"
repeat_region	10557..10964
	/rpt_family="ERV1"
repeat_region	10967..11463
	/rpt_family="L1"
repeat_region	12679..12741
	/rpt_family="MaLR"
misc_feature	13652..14115
	/note="similar to EST AI445193 (NID:g4287282) t191h10.x1"
repeat_region	16534..16822
	/rpt_family="Alu"
repeat_region	17605..17689
	/rpt_family="MIR"
misc_feature	18409..18912
	/note="similar to EST AA884467 (NID:g29993997) aj61c02.s1"
repeat_region	18461..18695
	/rpt_family="L2"
repeat_region	18866..19002
	/rpt_family="L2"
repeat_region	19033..19329
	/rpt_family="ERV1"
repeat_region	19780..19865
	/rpt_family="AchoBo"
repeat_region	19891..20438
	/rpt_family="L2"
repeat_region	20526..20678


```
QY 2003 ctatgtgttttgccttttcccttataataaataagacccttttccctccataaacttttgcatt 2062
Db 129299 CTAGTGTCTTACTTATTTCTCTTTATGATAATAA--CTTTTCCCTCAATAATTTTATT 129356
QY 2063 tctaagaaaataatagcacaatttcaaacataatgtttttacagtggtcttcttttt 2122
Db 129357 TCTAAGAAAAAT-----ATTCAACACTAATGCTTTACAAGTGT--TCATTTTC 129404
QY 2123 ttccocctgaagactaatgttgctcaataaaccacataagatttaagcatgagcagctg 2182
Db 129405 TTCCCTCAACAAAATCCAAATTTGGTCAGATAAACTACTAAGTGTGTGAGCGTGGACATC-- 129462
QY 2183 tigttagactacagattcagtttttttgatatattcttaattgtacttttggaatttta 2242
Db 129463 -TGTTAGAGTAGCAGATTCAGTGTGTTGGACACATTTTAAATGTGTACACTTCATCAATTTTA 129521
QY 2243 atttaagaagaagcaactg-----aaatigaaactctgagggcagctgactctacta 2293
Db 129522 ATTTAAGAAGGCCACCGAGATTGAAAGATTGAAATCTTGGGGCAGCTACATCTGTTA 129581
QY 2294 atgagccttattccattctctgatgttttttcaagaagaacactgcttgcattgattacgaa 2353
Db 129582 ATGAGCATTTGTCATTTTCATGATATATTTTAAAGAAATAA--ACTGCTTGAT--TACAGA 129637
QY 2354 tacactcagaagatcatttagcttgtagttggaattcttcttaagaaggaatgcttgaatt 2413
Db 129638 ATCATTTAAAAAGTACATAGCTGCTGCTGTTGAACTCTCATAAAGGAATTTTGAATTT 129697
QY 2414 ttctcattattgttttattgtttttatatacacttgccttattgaaatgtttagcagtatcc 2473
Db 129698 TTATAATATATTTCTTTTAAATACACTAGCT--CTTAGTTGAAAGTTTTGTGTACCC 129755
QY 2474 cttcccaacttatattgttgatattgatttgccttaggtaggagtttaaaaactttt 2533
Db 129756 CTTCTCCTCTTAAACTGTACTATGATGTGCTAGCCCAATCAACAGTTAAAAA--TTTT 129814
QY 2534 ccatggaactctgaacttaaacatacatgataacttaacttaactgtttaagaataa 2590
Db 129815 CCACATTAATGCTGCTGGTTAAACATACATATAA--TTATATAACTGCTAAAAAATGA 129870

RESULT 7
AC021847 AC021847 143812 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-181P4, WORKING DRAFT SEQUENCE,
DEFINITION 32 unordered pieces.
ACCESSION AC021847
VERSION AC021847.6 GI:8954243
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143812)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143812)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8567974.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0181P04
----- Summary Statistics -----
Sequencing vector: M13; 100%
```

```
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 125690 bases at least Q40
Consensus quality: 130641 bases at least Q30
Consensus quality: 133684 bases at least Q20
Insert size: 151000; agarose-gel
Insert size: 140712; sum-of-contigs
Quality coverage: 2.84 in Q20 bases; agarose-gel
Quality coverage: 3.15 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1345: contig of 1345 bp in length
* 1346 1445: gap of unknown length
* 1446 2816: contig of 1371 bp in length
* 2817 2916: gap of unknown length
* 2917 4030: contig of 1114 bp in length
* 4031 4131: gap of unknown length
* 4131 6612: contig of 2382 bp in length
* 6613 8567: contig of 1955 bp in length
* 8568 8667: gap of unknown length
* 8668 10333: contig of 1666 bp in length
* 10334 12876: contig of 2443 bp in length
* 12877 12976: gap of unknown length
* 12977 15219: contig of 2243 bp in length
* 15220 15320: contig of unknown length
* 15320 17854: contig of 2535 bp in length
* 17855 17954: gap of unknown length
* 17955 20526: contig of 2572 bp in length
* 20527 20626: gap of unknown length
* 20627 24354: contig of 3728 bp in length
* 24355 24454: gap of unknown length
* 24455 28206: contig of 3752 bp in length
* 28207 28306: gap of unknown length
* 28307 32673: contig of 4367 bp in length
* 32674 32773: gap of unknown length
* 32774 36150: contig of 3377 bp in length
* 36151 36250: gap of unknown length
* 36251 39810: contig of 3560 bp in length
* 39811 39910: gap of unknown length
* 39911 43789: contig of 3879 bp in length
* 43790 43889: gap of unknown length
* 43890 47197: contig of 3308 bp in length
* 47198 47297: gap of unknown length
* 47298 51028: contig of 3731 bp in length
* 51029 51128: gap of unknown length
* 51129 55156: gap of unknown length
* 55157 55257: gap of unknown length
* 55258 60672: contig of 5415 bp in length
* 60673 60772: gap of unknown length
* 60773 65411: contig of 4639 bp in length
* 65412 65510: gap of unknown length
* 65511 71071: contig of 5561 bp in length
* 71072 71171: gap of unknown length
* 71172 76415: contig of 5244 bp in length
* 76416 76516: gap of unknown length
* 76517 81312: contig of 4797 bp in length
* 81313 81413: gap of unknown length
* 81414 86835: contig of 5423 bp in length
* 86836 86935: gap of unknown length
* 86936 91556: contig of 4621 bp in length
* 91557 91657: gap of unknown length
* 91658 96939: contig of 5283 bp in length
* 96940
```



```

actin cytoskeleton"
/codon_start=1
/product="WASP-family protein"
/protein_id="BAA81795.1"
/db_xref="GI:5257244"
/translation="MPLVTRNIEPRHLCROFLPSVRSELECVTITLANVIRQLGSL
SYAEDIFGELETOANTASRVYSSLAERVDRLQVKTQDLPKEEVSQGIINTRKAFRS
KTIQDKLFDRNSLPVPVLETNYTFCDTPTPPLNLTPTVDDGKEALFYDPSYFDLW
KEMLOTKDLMKHKRKEKKNPNRNGVNPRIKTRKEWEMKMGOFVESKEK
LGTSGYPTLVONGSGTCVENDDASSYPPPOSASSPSSEONLPPPAEFSY
PVDNQRSGLAGPRSSVSVSHPPAPPLGSPGPKPGFAPPAPPPPPMIGIIP
PPPPVGGSPGTPPPSPSPSPHPDFAAPPPPPPPAAADYPTLPPPLPSQTGGAP
PPPPPPPPPPPTGADGQPAIPLPLSDTTKPKSLPAVSDLSLAIROGFOL
RRVEQRQRQRDRVGVNDVATILSRRLAVEYSDSEDDSDSEDDWDSD"
BASE COUNT 385 a 466 c 330 g 316 t
ORIGIN

Query Match 10.6%; Score 278.2; DB 85; Length 1497;
Best Local Similarity 64.9%; Pred. No. 5.3e-46;
Matches 428; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 243 atgcgcgtagtgaagaacatcagctatcctagcactgtgtgcacacagcagcctcctaga 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGCGGTAGTAACGAGAAACATCAGGCGACCTGTGCGGTACAGACGTGCTAGC 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 303 ggcattaagaatgaactgtgaacaaatattcttggcacaataatattagacaa 362
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 G---TTAGNAGCGAGCTGGAATGCTGTGACCAACATCACCCTGGCAATGTCTATCGACAG 117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 363 ctaagtgcctaaatgaatgtgaatattttggagaattattcaatgaagacacat 422
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 CTGGGCGAGCTGAGTAATATCAGAGGACATTTTGGAGAGCTCTTTACTCAGGCAAT 177
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 423 agttttcttcagagtcacactcatgtgaagaacgtgtggaccgtttatctgttagtgtt 482
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 ACCTTGGCTTCGCGGTAGCTCCCTTGTGAGAGGGTGCACCGACTACAGGTTAAAGTC 237
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 483 acacagcttgatccaaagaaagaagaattgttttgcagaatatacaaatgaggaagct 542
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 ACTCAGCTGGATCCCAAGGAAGAAGTGTCTACGCAAGGAATCAACACCCGAAAGCC 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 543 ttccgaagtctacaaattcagacagcagcgttttcgatcgaagactgtgctattccca 602
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 TTCAGAAAGTTCCACCATTCAGACAGAGAGCTTTTGTGACAGAAACTCTCTCCAGTGCCT 357
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 603 ttacagagacgtacagattgttgaacagcctcacctcctcaatatactcactccttat 662
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 GTCTTGAAGAACATACATACCTGTGATCTCTCCCTCTCAACAATCTTACCCTTAC 417
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 663 agagatgatgttaaagaaggtctgaagttttatataccaatccttcgtatttcttgatcta 722
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 AGGACGATGGAAGAGGAGCACTCAAAATCTACAGACCCCTTCATCTCTTTGATCTTT 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 723 tggaaagaaaaattgtcagaatagacagagagaagaaagaaagagagagagagag 782
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 TGAAGGAGAGAGAGTCTCAGGACACCAAGGATATCATGAAGAGAGAGAGAGAGAGAG 537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 793 cagaaaaatctagatcgtcctcatgaacacagaaaaagtgcgaagagcactcatgacagg 842
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 AAAAGAAAGAAAGAAATTAATCCAAATCGAGGGAATGTAAACCCACGCTAAATCAAGACAGT 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 843 cggcagaatgcagaagctggcccaaggtccagagctggctgaagatgatgctaactc 901
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 AAGGAAGAGTGGGAGAAATGAAGATGGGCAAGAAATTTGGAGTCCCAAGAAAGAGCT 656
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
HS393P12 104597 bp DNA PRI 23-NOV-1999
LOCUS HS393P12 Homo sapiens DNA sequence from PAC 393P12 on chromosome xpl1.21.
DEFINITION Contains a hypothetical protein KIAA0413 (KIAA0412, KIAA0065,
KIAA0569) LIKE Zinc Finger protein pseudogene, a hypothetical

```

Proline-rich protein KIAA0269 LIKE gene and a 60S Ribosomal Protein L7 LIKE pseudogene. Contains ESTs, an STS and a GSS (BAC end sequence), complete sequence.

AL022578
AL022578.1 GI:3080381
HTG; 60S Ribosomal Protein L7; KIAA0065; KIAA0269; KIAA0412; KIAA0413; KIAA0569; Zinc Finger human.

SOURCE
ORGANISM Homo sapiens
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 104597)
Bird, C.

REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Chromosome X Project Group (http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk

COMMENT
IMPORTANT: This sequence is the entire insert of clone 393P12. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX/
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The true left end of clone 393P12 is at 1 in this sequence. The true right end of clone 393P12 is at 104597.
393P12 is from the library RPC13 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

FEATURES
Location/Qualifiers
1..104597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.21"
/clone="RP3-393P12"
/clone_lib="RPC1-3"
35..331
/note="AluSx repeat: matches 1..297 of consensus"
990..1489
/note="MLTIF repeat: matches 541..15 of consensus"
2202..3095
/note="L1P15 repeat: matches 900..1 of consensus"
2949..3177
/note="L1 repeat: matches 5390..5160 of consensus"
3178..3565
/note="L1 repeat: matches 4374..4762 of consensus"
3622..4137
/note="L1 repeat: matches 4410..3874 of consensus"
4170..4468
/note="AluSx repeat: matches 299..1 of consensus"
4473..6201
/note="L1 repeat: matches 3840..2057 of consensus"
6295..6587
/note="AluSx repeat: matches 292..2 of consensus"
7254..7454
/note="L1P16 repeat: matches 904..690 of consensus"
7531..7631
/note="L1 repeat: matches 12..111 of consensus"
complement(7750..8788)
/note="match: cDNAs D16470 U00684 L23868; match: ESTs H11536 H06620"

Db 43058 AGTCACTCAGCTGGATCCCAAGGAAGAAGAGTGTCACTGCAAGTAATCAACACCCGAAA 43117

Qy 539 agcttcggaagtctcaaatcaagaccagcagcttttctgacgacgacgtttgcctat 598

Db 43118 AGCCTTCAGAGCTCCACATTCAGACACAGAGCTTTTGACACAAACTCTCTCCAGT 43177

Qy 599 tccattacaggagacgtacgtattgttgaaacagcctccacccctcctatactactcc 658

Db 43178 GCCTGCTCTAGAAACATACATACCTGTGATACTCTCCCTCTCAACAATCTTACCCC 43237

Qy 659 ttatagagatgatgtaagaagctgaagcttttataccactcctctgatttttga 718

Db 43238 TTACAGGACGATGAACAGAGGCACTCAAAATTCACACAGACCTTCTATCTTTGA 43297

Qy 719 tctatggaagaataatgttgcagatatacagagataagaggaggaagaaagaaqca 778

Db 43298 TCTTTTGAAGAAGAGATGCTGCAGACACCCAGGATATCATGAAGAAGAGAAAGCA 43357

Qy 779 gaagcagaaaaatctagatcgtctctcatgaaccagaaaaagtgccaaagcagcctcatga 838

Db 43358 CAGGAAGAAAGAAAGATAATCCAAATCGAGGGAATGTAAACCCACGTAATAATCAAGAC 43417

Qy 839 caggcggcgaagaatggcagaagctggcccaaggtccagagctggctgaagatgctctaa 898

Db 43418 ACGTAAGGAAGAGTGGGAGAAATTAAGATGGGGCAAGAAATTTGTGGAGTCCAAAGAAA 43477

Qy 899 tct 901

Db 43478 GCT 43480

RESULT 11

AF134304

LOCUS AF134304 1491 bp DNA PRI 14-JUN-1999

DEFINITION Homo.sapiens Scar2 (SCAR2) gene, partial cds.

ACCESSION AF134304

VERSION AF134304.2 GI:5053127

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 1491)

JOURNAL Bear, J.E., Rawls, J.F. and Saxe, C.L. III.

MEDLINE SCAR, a WASP-related protein, isolated as a suppressor of receptor

PUBMED defects in late Dictyostelium development

REFERENCE 98402540

2 (bases 1 to 1491)

AUTHORS Machesky, L.M. and Insall, R.H.

TITLE Direct Submission

JOURNAL Submitted (12-MAR-1999) Biochemistry, University of Birmingham,

REFERENCE Birmingham B15 2TT, England

3 (bases 1 to 1491)

AUTHORS Machesky, L.M. and Insall, R.H.

TITLE Direct Submission

JOURNAL Submitted (14-JUN-1999) Biochemistry, University of Birmingham,

REFERENCE Birmingham B15 2TT, England

REMARK Sequence update by submitter

COMMENT On Jun 14, 1999 this sequence version replaced gi:4927211.

FEATURES

Location/Qualifiers

source

1. 1491

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/map="Xp11.21"

<1..>1491

/gene="SCAR2"

/product="Scar2"

<1..>1491

/gene="SCAR2"

<1..>1491

mRNA

gene

CDS

/gene="SCAR2"

/function="may regulate actin dynamics through interaction with the Arp2/3 complex"

/note="WASP-family protein; similar to Dictyostellum suppressor of cyclic-AMP receptor"

/codon_start=1

/product="Scar2"

/protein_id="AAD33053.2"

/db_xref="GI:5053128"

/translation="PLVTRNIEPHLCROTLPVSRSELCVTNITLNVIRLGSLSK YAEIDFGEFLTQANTFASPVSSLAERVDGLQVKYQDLPKEEVSQVINTKRAFRSS KIQDLQKLFDRNLSKPKRKKEKKNRNGNPNRKRKTRKEWEKIKKQGEVESEKEL TKMLQDTRDLMKPRKKEKKNRNGNPNRKRKTRKEWEKIKKQGEVESEKEL VDNQYPPTLVQNGSIGCVENVDASSYPPSPQSDSSASPSFSEDNLPPPPAFESYP GFGNGSGLAGPKRSSVSPSHHPAPPLGSPSSKPGFAPPPAPPAPPMPMIGIIPP PPPIGFCSPCTPPPPSPSPFPDFAPPLPPPPAADYTLPPPLSOTRCAAPP PPPPPGPPPPPTFGAQCPAYPPPLSDTTKPKSLPAISDAHSDLSAICQGPYLRR VEEPEQEQKWDVVGVNDVATILSRRIAYEYSDSEDDSDSEFDGDDWSN"

BASE COUNT 391 a 462 c 319 g 319 t

ORIGIN

Query Match 10.0%; Score 262.4; DB 89; Length 1491;

Best Local Similarity 63.6%; Pred. No. 8.1e-43;

Matches 417; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

Qy 246 ccgctagtgaagaacacatcgatcctctagcgacactgtgtgccacacagcactgcctagagcg 305

Db 1 CCGTTAGTAAAGGAGACATCGAGCCAGGACCTGTGCTCGTCAGACCGTGCCTA--GC 57

Qy 306 attaagaatgaactggaatgtgttaaccaatatttctcttggcaaatataattagacaacta 365

Db 58 GTTAGAAGCGAGCTGGAAATGCTGACCAACATCACCTCGCAAAATGTCATCCGACAGCTG 117

Qy 366 agtgcctaaatgaatgtgtgaagataatttggaggaattattcaatgaagacacatagt 425

Db 118 GCGACCTTGAATTAATATGACAGAGGACATTTTGGAGAGCTCTTTTACTCAGCAAAATACC 177

Qy 426 ttctcctcagagtcacactcattgcaagacgtgtggaccgtttatctgtgtagtgtaca 485

Db 178 TTTGGCTCTCGGTAAGTCCCTTGTCTGAGAGGTCGATGACTACAAAGTTAAAGTCACT 237

Qy 486 cagctgtgtccaaaggaagaagaattgtcttcttgaagataatacaaatgaggaagctttc 545

Db 238 CAGCTGGATCCCAAGGAAGAAGAGTGTCACTGCAAGTAATCAACACCCGAAAGCCCTTC 297

Qy 546 cgaagttctacaattcaagaccagcagcttttcgacgacgacgttttcctattccatta 605

Db 298 AGAAGCTCCACCATTCAGAGACAGAAAGCTTTTTCAGAGAAACTCTCTCCAGTGCCTGC 357

Qy 606 caggagacgtacgattgttgaacagcctccactcctatactactcctcttataga 665

Db 358 TTAGAAACATACATACCTGTGATCTCTCCCTCTCAACAATCTTACCCCTTACAGG 417

Qy 666 gatgattgaagaaggctctgaagttttataccaatcctctgattttcttgatctatgg 725

Db 418 GACGATGGAACAGAGGCACTCAAAATTCACACAGACCTTCTATCTTCTTGTGATCTTGG 477

Qy 726 aagaagaaaaatgttgcagatacacagagataagaggaggaaggaaggaagcagaagcag 785

Db 478 AAGAAGAGAGTGTCTGACAGACACCCAGGGATATCATGAAGAAGAGAGAGAAACACAGGAAA 537

Qy 786 aaaaatctagatcgtcctcatgaaccagaaaaagtgccaaagcagcctcatgacagcgcg 845

Db 538 GAAGAAGAAAGATAATCCAAATTCGAGGGAATGTAAACCCACCGTAATAATCAACACAGCTAAG 597

Qy 846 cgagaatggcagaagctggcccaaggtccagagctggctgaagatgagtctaatct 901

Db 598 GAAGAGTGGGAGAAAAATTAAGATGGGGCAAGAAATTTGTGGAGTCCAAAGAAAGACT 653

RESULT 12

AF134305

[illegible]

```

Query Match          7.8%; Score 205.6; DB 85; Length 4450;
Best Local Similarity 64.6%; Pred. No. 2.5e-31;
Matches 323; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

QY 384 gctgaagatatatttggaagatttcaatgaagacacatagttttcttcagagtcacac 443
Db 1 GCTGAAGACATATTTGGTGGAGTCTGTTTAAATGAGGCTAACAACTTCTACATCAGAGCAAT 60

QY 444 tcaatgaagaacgtgtgagccgtttatctgttagttgattacacagcttgatccaaagaa 503
Db 61 TCTCTTCAAGACAGANTTGTATGTCGCTTCTGTCAAAGTCACCCAGCTGGATTCACACAGT 120

QY 504 gaagaattgtcttgaagatatatacaatgaggaaagctttccgaagttctacaaattcaa 563
Db 121 GAAGAGGTCCTCCTACATACAGATATCAACATGAAGAAAGCTTTCAAAGTTCCACAGTCCA 180

QY 564 gacacagagcttttcgacgacgaagcttgcctattcattacacaggagacgtacagatt 623
Db 181 GACCAGCAAGTGGTTTCAAGAACAGCAGATTCCTAATCCTGTTGCTGATATTTTACAACAG 240

QY 624 tgaagaacgctccacctcctcaataatactactcctcttatagatgatgtaagaaggt 683
Db 241 AGTGATAGCCACCCGCTCTGAACATCTGACACCATACAGATGACAAAGAGGATGGG 300

QY 684 ctgaagttttatcaacatcctctgattttcttctgattgattgaaagaaaaaattgtgcaa 743
Db 301 CTGAAGTTCTACTACTGATCTCTCTCTATTTCTTTGACCTCTGGAAGAGAAATGCTACAG 360

QY 744 gatacagaggaataagagaagaaaggaaggaagcagagcagagaaaaa---tctagatcgt 800
Db 361 GACACAGAAAGCAAAAGGAAAGAGAGAAAGGCGTCAAAAGGAGCAAAAGCGTATAGATGGC 420

QY 801 cctcatgaaccagaaaggtcccaagagcactcatgacagcgcgcgagagaatggcagaag 860
Db 421 ACCACCGCTGAGGTGAAAGAGGTGTAGAAAGGACCAAGAAACAGCGCCGAGGAGTGAATG 480

QY 861 ctggcccaaggtccagagct 880
Db 481 ATGGCATATGACAAAGACT 500

RESULT 14
AC021847/c
LOCUS AC021847 143812 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-181P4, WORKING DRAFT SEQUENCE,
32 unordered pieces.
ACCESSION AC021847
VERSION AC021847.6 GI:8954243
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
REFERENCE 1 (bases 1 to 143812)
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8567974.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0181P04
----- Summary Statistics -----

```

```

Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 125690 bases at least Q40
Consensus quality: 130641 bases at least Q30
Consensus quality: 133684 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 140712; sum-of-contigs
Quality coverage: 2.84 in Q20 bases; agarose-fp
Quality coverage: 3.15 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1345: contig of 1345 bp in length
* 1346: gap of unknown length
* 1446: contig of 1371 bp in length
* 2817: gap of unknown length
* 2916: contig of 1114 bp in length
* 4031: gap of unknown length
* 4131: contig of 2382 bp in length
* 6512: gap of unknown length
* 6513: contig of 1955 bp in length
* 8567: gap of unknown length
* 8668: contig of 1666 bp in length
* 10333: gap of unknown length
* 10434: contig of 2443 bp in length
* 12877: gap of unknown length
* 12976: contig of 2243 bp in length
* 15219: gap of unknown length
* 15320: contig of 2535 bp in length
* 17854: gap of unknown length
* 17955: contig of 2572 bp in length
* 20526: gap of unknown length
* 20627: contig of 3728 bp in length
* 24355: gap of unknown length
* 24455: contig of 3752 bp in length
* 28206: gap of unknown length
* 28307: contig of 4367 bp in length
* 32674: gap of unknown length
* 32773: contig of 3377 bp in length
* 36150: gap of unknown length
* 36251: contig of 3560 bp in length
* 39811: gap of unknown length
* 39911: contig of 3879 bp in length
* 43790: gap of unknown length
* 43899: contig of 3308 bp in length
* 47198: gap of unknown length
* 47298: contig of 3731 bp in length
* 51029: gap of unknown length
* 51129: contig of 4028 bp in length
* 55156: gap of unknown length
* 55257: contig of 5415 bp in length
* 60672: gap of unknown length
* 60772: contig of 4639 bp in length
* 65411: gap of unknown length
* 65511: contig of 5561 bp in length
* 71072: gap of unknown length
* 71172: contig of 5244 bp in length
* 76416: gap of unknown length
* 76516: contig of 4797 bp in length
* 81313: gap of unknown length
* 81413: contig of 5423 bp in length
* 86836: gap of unknown length
* 86936: contig of 4621 bp in length
* 91557: gap of unknown length

```

* * *	91557	96939:	contig of 5283 bp in length
* * *	96940	97039:	gap of unknown length
* * *	97040	103443:	contig of 6404 bp in length
* * *	103444	103543:	gap of unknown length
* * *	103544	109673:	contig of 6130 bp in length
* * *	109674	109773:	gap of unknown length
* * *	109774	118660:	contig of 8887 bp in length
* * *	118661	118760:	gap of unknown length
* * *	118761	130272:	contig of 11512 bp in length
* * *	130273	130372:	gap of unknown length
* * *	130373	143812:	contig of 13440 bp in length.
FEATURES			
source			
1..143812			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="6"			
/clone="rp11-181p4"			
1..1345			
misc_feature		/note="assembly_name:Contig10"	
	1446..2816		
misc_feature		/note="assembly_name:Contig12"	
	2917..4030		
misc_feature		/note="assembly_name:Contig14"	
	4131..6512		
misc_feature		/note="assembly_name:Contig15"	
	6613..8567		
misc_feature		/note="assembly_name:Contig17"	
	8668..10333		
misc_feature		/note="assembly_name:Contig18"	
	10434..12876		
misc_feature		/note="assembly_name:Contig19"	
	12977..15219		
misc_feature		/note="assembly_name:Contig20"	
	15320..17854		
misc_feature		/note="assembly_name:Contig21"	
	17955..20526		
misc_feature		/note="assembly_name:Contig22"	
	20627..24354		
misc_feature		/note="assembly_name:Contig23"	
	24455..28206		
misc_feature		/note="assembly_name:Contig24	
	clone_end:SP6		
	vector_side:right"		
misc_feature		28307..32673	
	/note="assembly_name:Contig25"		
misc_feature		32774..36150	
	/note="assembly_name:Contig26"		
misc_feature		36251..39810	
	/note="assembly_name:Contig27"		
misc_feature		39911..43789	
	/note="assembly_name:Contig28"		
misc_feature		43890..47197	
	/note="assembly_name:Contig29"		
misc_feature		47298..51028	
	/note="assembly_name:Contig30"		
misc_feature		51129..55156	
	/note="assembly_name:Contig31"		
misc_feature		55257..60671	
	/note="assembly_name:Contig32"		
misc_feature		60772..65410	
	/note="assembly_name:Contig33"		
misc_feature		65511..71071	
	/note="assembly_name:Contig34"		
misc_feature		71172..76415	
	/note="assembly_name:Contig35"		
misc_feature		76516..81312	
	/note="assembly_name:Contig36"		
misc_feature		81413..86835	
	/note="assembly_name:Contig37"		
misc_feature		86936..91556	
	/note="assembly_name:Contig38"		
misc_feature		91657..96939	
	/note="assembly_name:Contig39"		

